DOCUMENT RESUME

ED 349 814 FL 020 633

AUTHOR	Oller, John W., Jr.
TITLE	Language Testing Research: Lessons Applied to LEP Students and Programs.
PUB DATE	Aug 92
NOTE	93p.; In: Focus on Evaluation and Measurement. Volumes 1 and 2. Proceedings of the National Research Symposium on Limited English Proficient Student Issues (2nd, Washington, DC, September 4-6, 1991); see FL 020 630.
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Intelligence; *Language Proficiency; *Language Research; Language Tests; *Limited English Speaking; *Testing

ABSTRACT

This report on language testing research focuses on lessons applied to limited English proficient (LEP) students and programs. First, a review of the history of primary and non-primary language testing is provided. The question of why there is no field of primary language testing is then discussed. The second major section of the report is a review of the broader literature of educational measurement as it relates to the critical role of language proficiency. The third section offers an idea of the place of language proficiency in a broader theory of human intelligence and representational capacities. Building on findings in non-primary language research, a possible resolution of the apparent controversy over the old notion of a single unifying general intelligence and distinct multiple intelligences is proposed. To conclude, a few observations about how to go about testing the increasing number of LEP students in schools are presented. Deep surface rather than surface assessment through discourse-based, real-life performances are recommended. Two responses to the paper, one by Fred Davidson and the other by Myriam Met, are appended. (VWL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Language Testing Research: Lessons Applied to LEP Students and Programs

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

John W. Oller, Jr.
University of New Mexico

ED349814

> ...man is not just a creature of accident, chained to and formed by the particular cave in which he is born....No real teacher can doubt that his task is to assist his pupil to fulfill human nature against all the deforming forces of convention and prejudice... Moreover there is no real teacher who in practice does not believe in the existence of the soul, or in a magic that acts on it through speech (Allan Bloom, 1987, *The closing of the American mind: How higher education has failed democracy and impoverished the souls of today's students*, p. 20).

For educators at large, probably the first and most important lesson learned from language testing research is that *language proficiency (whether it is construed as a general factor or as a constellation of related abilities) is important in one way or another to nearly everything that takes place in education -- whether at school or elsewhere.* Language proficiency is a critical element in the process of becoming literate and all of the other public manifestations of human intelligence that enable us to be the social beings that we are. It is important to intrapersonal and interpersonal performances of all sorts. Language, perhaps more than any other aspect of our existence, is what enables us to be members of a community that includes people other than ourselves. Perhaps I can be forgiven, as someone who comes partly from a foreign language teaching background, for stressing as enthusiastically as I do that proficiency in another language is like a key that opens a door to new worlds of understanding and provides access to new communities. However, if we remain in a permanent state of *monolingual myopia* -- which in its most pernicious form is a terminal disease -- language can be a wall that separates us from all the world beyond our particular primary language community. To the *terminally monolingual*, the wall is invisible, intangible, and seemingly non-existent. Yet is it as impenetrable as solid granite and forms a prison more secure than concrete and steel ever could. Electronic surveillance in the prison is altogether unnecessary because the inmates are as unaware of their situation as Plato's inhabitants of the cave were of theirs.

FL 020 633

The good news, of course, is that by acquiring a language or two beyond our primary linguistic system, we can become more aware of our limitations, prejudices, and the inevitable ignorance that plagues all the denizens of all the caves, and to some extent, we can, it seems, escape the special prison of monolingual prejudice. With this desir-

2 BEST COPY AVAILABLE

able aim in mind, the insight that I want to develop—that *language proficiency is central to all aspects of education* -- if it can be called an insight, will be news to no one in the bilingual education arena. Nor is it apt to make headlines with teachers who work with students of limited English proficiency (LEPs). Still, it is an insight that bears scrutiny and certainly criticism, and it epitomizes, I believe, what language testing research has to offer to a conference on evaluation and measurement issues relative to LEP students and the programs that aim to serve them. With respect to the evaluation of programs, a special sort of assessment problem, I concur with Prestine (1990) where she cites Rist (1982) who notes that program evaluation inevitably entails a general question that "is at once disarmingly simple and incredibly complex" -- namely, "What's going on here?" (Rist, 1982, p. 440, and Prestine, 1990, p. 288). I'll try to show that language proficiency is a critical element in answering this general question not only in relation to individual students but also with respect to program evaluation.

For the particular group of educators assembled at such a conference as this one, I doubt it will be necessary to sell the idea that *language proficiency matters*. This is something that I assume we all agree on from the start. We may differ, however, in subtle and unanticipated ways on just how language proficiency matters and to what degree it matters. What I will attempt to do, therefore, is to elaborate on the ways in which language proficiency seems to matter according to the evidences afforded by theory and research. My analysis will be based on a selective review of the relevant literature. Underlying all of the discussion will be the ultimate aim of reaching some practical conclusions concerning how we ought to go about testing and evaluation in educational programs for LEP students. The best I can hope for is to affirm some of the good things that are already happening, to offer some constructive (I hope) criticisms concerning theories and practices that need mending, and to encourage us to capitalize still more on the rich linguistic resources that are coming to us in ever greater quantities from a pluralistic world of many languages.

To that end, I would like to suggest that the first corollary of my starting premise, that *language proficiency is a central element of all educational undertakings*, might be that the term "limited-English-proficiency" implies a complement of "almost-unlimited-proficiency-in-some-other-language-or-languages." While I do not want to deny the benefits (or importance) of students acquiring a high degree of proficiency in English in these United States, I do want to suggest that it is strange that our educational systems and national policies (as diverse and amorphous as they may be; see Prestine, 1990, for a discussion of great interest) seem generally determined (at least in practice) to either ignore or to deliberately remove rather than to nurture and preserve the linguistic resources that are literally walking into our schools at an ever increasing rate. Corresponding to the common emphasis on limitations, disabilities, disorders, disablements, disenfranchisements, etc., it seems to me that there ought to be greater consideration of the positive complements of these terms. In this suggestion, I concur with Lynda Miller (1990) where she contrasts her emphasis on "competencies" (taking her cue from the term "multiple intelligences" as employed by Gardner, 1983 and seq.) with the more common "approach in which the emphasis is on deficits and disabilities" (p. 2) or on "impairments, handicaps, and disorders" (p. 4).

According to the positive complement of the "deficit approaches" -- which might be properly called "empowerment approaches" -- the attainment of language proficiency is perhaps the main road to social empowerment (Cummins, 1986). As Miller puts it (following Hirsch, 1987): "being literate...is possessing shared background knowledge and holding positions of responsibility and power at the macro-levels of society" (Miller, 1990, p. 3). David Olson (1986) goes so far as to suggest that intelligence itself is hardly more than "literate competence" (p. 338) or "the distinctive forms of symbolic systems evolved and exploited by a culture as a means for representing and acting on the world" (p. 345).[1] Even Walters and Gardner (1985) who think in terms of "multiple intelligences"; also see Gardner, 1983, 1989, 1990) say that in their later development "children demonstrate their abilities in the various intelligences through their grasp of various symbol systems" (p. 15). In fact, each separate intelligence, of the seven they advocate (which we review in part 3, below), is eventually seen "through a *symbol system*: language is encountered through sentences and stories, music through songs, spatial understanding through drawings, bodily-kinesthetic through gesture or dance, and so on" (p. 15).

These ideas, though not identical with the view that I would like to advocate and develop here, still point, as I understand them, in the direction we ought to follow, and all of them tend to show the central importance of symbolic systems of which, I will endeavor to show (following C.S. Pierce [1839-1914]), natural language systems are chief. At any rate, all of the foregoing provides, I hope, a suitable preamble, a jumping off place, for the development of my main argument which follows in four parts which I will preview immediately.

I begin with (1) a review of the history of primary and non-primary language testing and with a provocative question: how come there is no field of primary language testing? This quandary, will be resolved early in the discussion in a way that illustrates my starting point above about monolingual myopia. It turns out that there are in fact many approaches to the measurement and testing of primary language skills, but that nearly all of them have been mis-identified as pertaining primarily to some other actually incidental purpose. This was unlikely to be noticed, however, owing to the pervasive monolingual myopia that has been prevalent for more than a century of public schooling and that still pervades the American educational scene. Until research on the testing of non-primary language proficiency began to bud in the late 1950s, hardly anyone ever thought to ask about research into the character of primary language proficiency. For this reason, the ideas to be gleaned from non-primary language testing especially, may be of some use to educators at large as well as those who work with the growing numbers of LEPs in our schools.

In order to see the connections of research in non-primary language measurement with broader issues in education, the second major section of this paper is a review of (2) the broader literature of educational measurement as it relates to the central theme -- the critical role of language proficiency. We will view that theme from a variety of angles and try to develop an up-to-date idea of where we are at present with respect to the unwieldy problem of measuring LEP students and evaluating the programs that purport to serve them.

The third major section of the paper offers (3) a somewhat elaborated idea of the place of language proficiency in a broader theory of human intelligence and representational capacities. Along the way, I will try to point out general themes of agreement and certain contrasting trends, e.g., the traditional views of general intelligence as contrasted with multiple intelligences as proposed back in the 1930s by L.L. Thurstone and others and revived and invigorated in recent years by Howard Gardner, Joseph Walters, Vera John-Steiner (1985), and others. Building on findings in non-primary language testing research, I will propose a possible resolution of the apparent controversy over the old notion of a single unifying general intelligence and distinct multiple intelligences. I will argue that these theories are not incompatible, but rather that they are complementary ways of viewing different facets of distinctive human abilities.

Finally, I will conclude with (4) a few observations about how we might go about the practical business of testing (and also of teaching) the increasing numbers of LEP students that are working their way through our schools. I will recommend deep rather than surface assessment through discourse-based, real life performances.

(1) Research in Primary and Non-Primary Language Testing

In undertaking a review of research on language testing, as soon as we begin to talk about "non-primary language testing" we are bound to ask: Why is there no distinct field of primary language testing? The answer to this question is that many approaches to the business of measuring primary language skills do in fact exist, but that they go by many different names. For instance, "intelligence testing" generally aims at primary language proficiencies and "verbal intelligence testing" specifically does so. Measures of listening and speaking abilities, speech and hearing tests, literacy tests of all sorts, but especially tests of reading vocabulary, reading comprehension, and writing proficiency tests clearly aim at primary language skills. In addition to the traditional categories of intelligence and achievement tests, there are many deficit oriented categories of primary language assessment: e.g., tests of "language disorders," "learning disabilities," "mental retardation," and more recently many different sorts of "cognitive" and "metacognitive" tests, not to mention "linguistic" tests, "sociolinguistic elicitation devices," tests aimed at "discourse abilities," "grammatical intuitions," "metalinguistic awareness," etc. I submit that there are many reasons why these various approaches to primary language assessment have not been recognized as a coherent branch of educational measurement, but none, I suppose, is more important than the general affliction of American educators with what I am calling here, monolingual myopia. I hasten to add that I am not saying that there are no important differences among the various fields of study listed in this paragraph, nor am I suggesting that primary language proficiency is the only object of interest. What I am saying is that all of the foregoing measurement efforts, and many others that I have not named, have as their principal, unstated object, the measurement of one or another aspect of primary language ability.

Hakuta (1986) has done an excellent job of illustrating the misclassification of many immigrants to the United States ever since the early decades of the twentieth century. He traces deficit theories of bilingualism back to fallacious interpretations of "IQ" tests that were actually little more than measures of English proficiency. More recently, Gardner and Hatch (1989) observe that "linguistic and logical-mathematical symbolization" predominate in both the curriculum and the school tests of "achievement, aptitude, and intelligence"(p. 6). This same complaint against traditional approaches to the study of intelligence in particular is what has led Gardner (1983, 1989, 1990) and his collaborators (also see Walters and Gardner, 1985, 1986a, 1986b) to develop the theory of "multiple intelligences". However, I submit that if it was the prevalence of monolingualism among the American educators that held the reigns of power from the early part of this century that set them up to misinterpret a mere lack of proficiency in English as a second language as a widespread intelligence deficit among children and adults from non-English speaking backgrounds. As Hakuta (1986) shows, immigrants in the early decades of the twentieth century were often described as "linguistically confused," "mentally retarded," "learning disabled," and so forth. By now it is clear that measures of yet to be acquired language skills were simply misidentified as indicating deficient cognitive powers of a much deeper sort.

Moreover, as Ortiz and Yates (1983) have shown, the problem is far from solved as we approach the twenty-first century. In Texas alone, as recently as eight years ago, Ortiz and Yates found that Hispanics were grossly over-represented (about 300 percent) in classes for the mentally retarded and other exceptionalities. Interestingly, as Cummins (1984) points out, the American Association of Mental Deficiency still depends on IQ scores (formerly one but now two standard deviations below the mean) as a part of its definition of "mental retardation" (McKnight, 1982). But why should anyone expect Hispanic children to have a 300 percent higher incidence of mental retardation than other ethnic groups in Texas? What most of those Hispanic children obviously have in common is Spanish rather than English as their first language. A small percentage of them, probably no greater than the percentage in other ethnic groups, may have some form of genuine mental deficiency, but there is every reason to suppose that the vast majority of Hispanic children in Texas are quite normal in their general mental abilities.[2] Because so many of them, however, have been misidentified as exceptional we may suppose that some children with genuine difficulties have also been overlooked and are not getting the special educational they need.

At least since the time of Francis Galton [1822-1911] (see Galton, 1869) -- Darwin's cousin and precursor of the modern intelligence testing movement -- which is generally credited to Alfred Binet [1857-1911] (see Binet and Simon, 1905) language proficiency tests have often been misinterpreted as measures of something else. For instance, Binet himself wrote:

> One of the clearest signs of awakening intelligence among young children is their understanding of spoken language...(1911, p. 186).

He said that according to teachers of his day the best way to form an impression of a child's intellect was to "talk to him" (1911, p. 308). In fact, the Binet and Simon (1905) tests included such obvious language proficiency tasks as responding to commands (e.g., "Point to your nose"), repeating a phrase or sentence, naming objects, telling what's going on in a photograph, answering simple questions (e.g., "What's your name?" "Are you a boy or a girl?" etc.), counting coins, copying a phrase or sentence, reading aloud and recalling points of information, writing phrases from dictation, defining words, etc.. All of this is relatively harmless so long as the language of the testing is the child's primary language system, but when it is not, difficulties arise. The nearly complete confounding of language proficiency with native intelligence persisted in the thinking of Binet who seemed to vacillate between the view that intelligence was distinct from acquired skills (Binet and Simon, 1905, p. 42) or that it was something that developed with "instruction" (p. 289). In the year of his death he wrote that children of higher standing manifest their "intellectual superiority" mainly "in tests where language plays a part" (p. 321).

The confounding of language proficiency with innate intelligence was especially apparent in a variety of fill-in-the-blank (cloze procedure) used by the German psychologist, Hermann Ebbinghaus [1850-1909]. According to David Harris (1985), as early as 1897, Ebbinghaus applied cloze procedure (more than half a century before its formal christening by Wilson Taylor, 1953) to meaningful prose with the intent of measuring the intelligence of school children. In the venerable tradition of *Gestalt* psychology, Ebbinghaus contended that intelligence involved linking elements so as to form coherent wholes. As paraphrased by Whipple (1915), Ebbinghaus is reported to have said:

> To measure intelligence, therefore, we must employ a test that demands ability to combine fragments or isolated sections into a meaningful whole. Such a test [that he called *Kombinationsmethode*] may be afforded by mutilated prose, i.e., by eliding letters, syllables, words, or even phrases, from a prose passage and requiring the examinee to restore the passage, if not to its exact original form, at least to a satisfactory equivalent of it (p. 285; also quoted in Harris, 1985, p. 367).

Marion Rex Trabue, about 1914 according to Harris, claimed to have improved the procedure by applying it to isolated sentences. Trabue argued that using isolated sentences, rather than connected prose, allowed him to rank items by difficulty thus creating a near interval scale and giving higher reliability in scoring. While Trabue's insistence on using disconnected sentences was, in my estimation, a step backward from where Ebbinghaus began, Trabue was among the first to explicitly say that his tests were measuring "language ability" (Trabue, 1916, p. 1). In spite of this, Trabue-type fill-in tasks based on isolated sentences continued long afterward to be applied in so-called "intelligence" tests which were supposed to be measures, not of acquired language skills, but of innate abilities (e.g., tests by E. L. Thorndike, Lewis M. Terman, and others).

Subsequently the various tasks recommended by Binet and others were reinterpreted, and alternately amplified and reduced several times, and were eventually canonized into various modern IQ tests (Binet and Simon, 1905; Terman, 1925; Terman and Oden, 1947; Terman and Merril, 1960; Kaufman, 1979). The best known examples of IQ tests are divisible roughly into the categories verbal and non-verbal (or performance) tests. In the non-verbal category *Raven's Progressive Matrices* and *Cattell's Culture Fair Test of Intelligence* are often used. Batteries aimed at both categories, however, are also well known: e.g., the *Thorndike-Lorge*, the *WISC-R*, the *Otis-Lennon Test of Mental Abilities*, etc.

Arthur Jensen of UC Berkeley fame (cf. Jensen, 1969, 1980) and Richard Herrnstein (1973; also Herrnstein and Wagner, 1981) of Harvard, extended the IQ testing movement, it would seem, to its

8

most extreme limits by claiming to be able not only to reliably determine innate intellectual capacities but to distinguish races and ethnic groups according to such measures. Most thinking persons find their reasoning spurious and their claims unconscionable—a kind of intellectual atavism harking back to racist theories of the philosopher Nietzsche and the idea of an intellectual aristocracy promoted in relation to the eugenics movement that began with Sir Francis Galton (1869). While such views have been severely criticized (and, I believe, properly so; see Mercer, 1973, 1984; and Gould, 1981), the best argument against them has largely been overlooked: namely that what the traditional intelligence tests measure best are acquired primary language skills. This idea is latent in the recent literature on "multiple intelligences," but has rarely been brought to bear as some believe it should (Oller, 1991). For instance, Walters and Gardner (1985) say, "We speculate that the usual correlations among subtests of IQ tests come about because all of these tasks in fact measure the ability to respond rapidly to items of a logical-mathematical or linguistic sort" (pp. 13-14). This very nearly amounts to saying that what those tests mainly measure is primary language proficiency (Oller and Perkins, 1978).

In spite of the long history of primary language testing from the early 1900s forward under the guise of IQ measurement, the notion of language proficiency per se, would progress little until empirical studies of foreign language proficiency began to appear in the late 1950s. Among the first was Carroll, Carton, and Wilds (1959) showing that cloze procedure had some potential as measures of language proficiency. A spate of studies would soon follow (Carroll, 1961; Lado, 1961; Valette, 1964; 1967) but it would not be until that latter part of the 1960s that non-primary language testing research would begin to flourish (cf. Upshur, 1967; Upshur and Fata, 1968; Spolsky, 1968a, 1968b, Anderson, 1969; Upshur, 1969a, 1969b; Oller, 1970; Oller and Conrad, 1971; Savignon, 1971). From there forward, too many research reports, conferences, and books would be generated for them to be adequately covered in any single review. However, it would not be until June, 1984 that the first issue of the journal *Language Testing* would appear. By then certain general themes and trends had been fairly well defined and the many of the paths that are currently being followed out had been marked off. Rather than try to plod through the whole terrain, in what follows I will concentrate on what I think the most important themes were in the 1970s and 1980s and still are in the 1990s.

It was John Carroll (1961) who suggested the distinction between discrete point approaches and integrative approaches to language testing. Discrete-point tests were grounded in the taxonomic approaches to linguistics that would later fall into disfavor as the Chomskyan revolution (see Chomsky, 1956, 1957, 1965, 1972, 1975, 1980a, 1980b, 1988) began to have its fuller impact into the 1970s and 1980s (see Newmeyer, 1980). Discrete point tests were based on inventories (taxonomies) of various sorts of elements. For instance, the phonological system of a language was supposed to consist of phonemes which could be tested one by one. The lexicon was a list of words, and grammar (alias syntax) was a list of patterns. This taxonomic way of looking at language, and at human abilities in general, still prevails among many (though certainly not all) psychologists (cf. the numerous examples cited by Cummins, 1984), speech-language pathologists (Coles, 1978, and Cummins, 1986, document this claim), and educators in general (Cummins, 1984, 1986; Cummins and Swain, 1986; Bloom, 1976; Bloom and Krathwohl, 1977; Swanson, 1988).

According to the discrete-point model, a sufficient number of items aimed at elements drawn from the several inventories of phonemes, morphemes, lexical items, and syntactic patterns would assure a valid test of language proficiency. In the 1980s, this same taxonomical thinking would persist in lists of "notions" and "functions" of speech acts and discourse (cf. Farhady, 1983b, and his references). The latter extension was certainly a natural one, but it did not really depart from discrete-point theory. The purest varieties of such thinking, e.g., Lado (1961) contended that language test items should focus on only one skill (e.g., listening), and only one domain (e.g., phonology), and only one element (e.g., a particular phonemic contrast) at a time. Besides distinguishing domains of structure -- phonology, morphology, lexicon, and syntax (semantics and pragmatics were not much thought of during the discrete-point heyday) -- discrete-point testers also distinguished skills (listening, speaking, reading, and writing). It was claimed that a test item could not be very good if it mixed several skills and/or domains of structure. And this contention itself pointed to what Carroll (1961) called "integrative tests."

For instance, Robert Lado (1961) contended that giving dictation, a foreign language testing technique popular with language teachers (cf. Valette, 1964; Finocchiaro, 1964), was not a good method because it mixed everything together. It was integrative rather than discrete-point (i.e., taxonomical) in its orientation. According to Lado, dictation did not test phonemic contrasts since these were apt to be given away by lexical or syntactic context. It did not test words because the words were "given" by the person reciting the material to be written down. It did not test syntax since the syntax also was "given." Worse yet, according to discrete-point thinking, dictation mingled listening comprehension with writing and reading. It also mixed phonology, vocabulary, morphology, and syntax (not to mention semantics and pragmatics) into a potpourri.

Discrete-point theory, however, in the final analysis was more of a hypothetical perspective than a practical one. Had it been influenced much by empirical evidence, it would have had to be radically

BEST COPY AVAILABLE

revised since language students in taking dictation do make many errors in just the domains that Lado claimed were not tested. For instance, in actual dictation protocols, we find evidence of phonemic contrasts that have been obliterated, for example, "collect" is apt to be rendered "correct" by an Asian writing a dictation in English. Or, complex consonant clusters of certain types of morphological inflections are apt to be omitted in many cases. Furthermore, the same persons who make these sorts of errors in taking dictation are apt to make analogous errors in writing an essay, speaking, or other discourse processing tasks. In fact, such problems carry over into relatively routine tasks such as repeating sequences of heard material, reading aloud, or even copying a text.

Also, in taking dictation, word order is sometimes adjusted in surprisingly creative and ungrammatical ways. Lexical items are changed radically. For example, in one study at UCLA a passage on "brain cells" was rendered in an almost coherent way by one non-native speaker of English as a text on "brand sales." Almost everything in the text was changed though a superficial phonetic resemblance remained between what had been dictated and what was written down. Less dramatic transformations of the same sort are commonly observed in dictation protocols (cf. Oller, 1979, pp. 283-285, for several examples).

As I argued in 1979 (p. 266) and continue to believe today, discrete-item tests do not accord well with what people do when they process text or discourse in normal ways. An example of a test exemplifying early discrete-point, taxonomical theory that has been widely applied but without much success is the Carroll and Sapon (1959a) *Modern Language Aptitude Test* (also see their *Manual*, 1959b). Carroll (1967) found, in a massive study of college foreign language majors near graduation, that the *MLAT* was only a significant predictor of foreign language attainment if extraneous variables such as interest, parental language background, and travel to the foreign country were included in the regression equations. Even with these extraneous variables added in, the *MLAT* still accounted for a modest 9 percent or less of the total variance in foreign language attainment. The several subtests of the *MLAT* itself, however, accounted for less than 1 percent of the total variance in foreign language attainment. More recently, Goodman, Freed and McManus (1990) again found the *MLAT* to be a non-significant predictor of success in foreign language courses for 586 students tested at the University of Pennsylvania. They speculated that perhaps the failure of the *MLAT* in this case was due to the fact that language teaching seems to be moving more and more in the direction of integrative, whole language approaches.

It is possible to find many examples of integrative tests that actually proved more robust both in theory and in practice than discrete-item tests. These included dictation (Valette, 1964), essays (Briere, 1966), answering questions orally (Upshur, 1967, 1969a), telling a story (Politzer, Hoover, and Brown, 1974), giving a speech, conversation or oral interview (ETS, 1970), reading aloud (Kolers, 1968), answering questions about a text (Politzer, Hoover, and Brown, 1974), repeating sequences from a text or narrative (also known as "elicited imitation"; Baratz, 1969; Politzer, Hoover, and Brown, 1974; Swain, Dumas, and Naiman, 1974), translating from L1 to L2 or the reverse ("elicited translation"; Swain, Dumas, and Naiman, 1974), etc. One of the various integrative types of task experimented with in the late 1960s and early 1970s was cloze procedure -- a method christened as such by Wilson Taylor (1953, 1956, 1957) for measuring readability of texts. It involves omitting words from a written (or possibly oral text) and requiring the examinee to replace the missing items (Anderson, 1969; Spolsky, 1968; Oller and Conrad, 1971; Oller, 1973).

As empirical research began to accumulate in the 1970s and into the 1980s it became clear that there were practical as well as theoretical differences between integrative and discrete-point tests. Integrative tests were apparently measuring some traits and abilities of language users that discrete-point tests could not get at. Still, even into the 1970s there were some, Earl Rand of UCLA, for instance, who insisted that discrete-point methods were either better or at worst equivalent to integrative tests (Rand, 1972, 1976). These claims were rarely sustained in practice. If one had examined closely the empirical results, it would have become clear that greater reliability and greater validity generally accrued to tests falling toward the integrative end of the spectrum.

Farhady (1983a) disagreed with this claim, but his examples were, as Oller (1983b, p. 321 footnote a) pointed out, drawn from tests that were quite integrative in character. Therefore, when Farhady (1983a) claimed that there was no difference between integrative and discrete-point tests with respect either to reliability or validity, he was really saying in effect that there is little difference between several about equally integrative tests. He was comparing reasonably good oranges with other reasonably good oranges. There were no truly discrete item tests in the inventory he compared. In any event, it is illogical to argue that the kind of test item that fully isolates a particular phonemic contrast, or a single lexical item, or a particular grammatical morpheme, or a syntactic rule, will yield results equivalent to the sort of test that requires the employment of a vast system of such relationships -- a whole grammar. If those two types of tests did turn out to be equivalent (which they are not, see also Damico and Oller, 1980; and Damico, Oller, and Storey, 1983), the result would be entirely anomalous as there simply is no theory whatever that predicts such an outcome. If a given phonemic contrast, say, /r/ versus /l/, is not in some sense distinct from, say, the syntactic transformation that copies the number of a referring head noun onto its respective present tense verb and its demonstrative modifier, e.g., in "These recommendations are...", then the distinction between phonology and syntax must be misguided. But how? While tests of particular phonemic contrasts, or inflectional morphemes, or syntactic rules, might generate reliabilities in the range of .6 to .7 (e.g., Evola, Mamer, and Lentz, 1980), tests of a more integrative character generally yield reliabilities about 10 points higher in the range of .8 to .9 (Oller, 1972, for instance). Or consider the fourteen different integrative tasks used in research to calibrate the language question on the 1980 U. S. Census, none yielded a reliability lower than .98 (cf. Scott, 1979).

It seemed to many, therefore, toward the end of the 1970s that integrative testing had prevailed over discrete-point approaches. However, this conclusion may have been premature. In the context of normal language processing, any given discrete-point item of interest may always be singled out for special attention *in that context*. On the other hand, a single element of any sort (a thoroughly isolated discrete-point) in the absence of the dynamic tensional context of discourse is like the sound of one hand clapping. Such discrete-points become mere fictions, like the dimensionless points of a line. Without the line, the points along it are dimensionless locations occupying space exactly nowhere. In context notions of discrete elements of language structure or skill are valuable theoretical constructs, but without the context, they are undefined fictions.

Out of the controversy over discrete-point versus integrative tests, there emerged a distinction of a different sort. While the original dichotomy (proposed by Carroll, 1961) was based on superficial aspects of test items, domains of structure, and modalities of processing, it became increasingly clear that the distinction had been incompletely and inadequately drawn. Carroll (1961), Rand (1976), and Farhady (1983a) all observed that there never was a truly categorical difference between discrete-point and integrative test items. The difference was merely one of degree. The dichotomy formed a continuum whose end-points were fully distinct only in theory. In practice, there are no completely discrete-point tests anymore than there are points or lines in the space/time continuum apart from some object or trajectory to define them. In actual experience all test items are more or less integrative in character.

Normal language use always involves meaning beyond the theoretically discrete elements of surface forms. That is, there is a linking with persons, places, things, events, relations, etc., in experience. However, if this meaning aspect beyond surface form is admitted, no test item can meet the demands of discrete-point theory. As I have hinted several times above, it may be worth saying straight out at this point that semantics and pragmatics were notably absent from discussions of discrete-point items. This was probably due to the fact that meaning as such is never a discrete-point affair. It cannot be since meaning spills over into the whole continuum of experience which the very existence of meaning both presupposes and implies.

Another insurmountable difficulty for discrete-point theory was that language use occurs in real time and is therefore time-constrained. This is not so obviously true for reading and writing as it is for listening and speaking tasks. However, it is easy to prove with a little thinking that in fact there are severe temporal constraints on reading and writing as well as on oral tasks. Meanings that involve long-range constraints in a written text, for instance, are essentially inaccessible to persons who lack a certain level of language proficiency owing to the limited time that they can hold the target language material in working memory. If the requisite part of the memory image fades from consciousness before the part with which it must be linked can be grasped, it will be impossible because of this temporal fact to grasp the full meaning.

Moreover, there are many other ways that real time constraints operate with reference to reading and writing in respects that are precisely analogous to temporal constraints on oral tasks. For instance, we may not have time to go and ask someone what So-and-So's last name is so we can look him up in the phone book. Or, we may not have time to drive to the library to look up a particular reference for a research paper. We may spend hours looking for a certain statement in a large book, or several volumes. These cases are hardly different from the problem of trying to recall some significant detail from a conversation (e.g., did he say to turn right or left on Oak Street?). In the final analysis, the salient differences between speech and writing seem less so when we look more closely at each one. Time and meaning, respectively, constituted the pragmatic naturalness constraints that led to a differentiation, therefore, of a certain subclass of integrative tests that came to be known as pragmatic (Oller, 1973, 1979; Cohen, 1980; Savignon, 1983). This subclass, it turned out, was entirely distinct from discrete-point tests. In fact, the pragmatic naturalness criteria eliminate any strictly discrete-point item as unnatural. Such items do not really involve normal language use anymore than the recitation of a number or parroting a numerical operation constitutes mathematical reasoning.

In addition, many tests that are thoroughly integrative in character also fail to meet the pragmatic naturalness criteria. For instance, the proofreading test explored by Barrett (1976) was integrative but failed the meaning criterion. It involved the omission of morphologically redundant elements (e.g., plural markers, tense indicators, articles, prepositions, verb particles, etc.) from prose and required the restoration of these elements by examinees. A peculiarity of the task was that fluent readers had to attend so much to the surface form of the text in order to notice the missing elements that they failed to process the meaning of the text and after performing the task could not even tell what the text was about. On the other hand, examinees who did concentrate on the meaning, and who could answer reasonable questions about its content, would invariably get low scores. These results are consistent with the frequent observation by proofreaders that plying their trade slows down their reading. In fact, they often resort to rather unusual methods of checking surface forms such as reading the text backwards, or following it word-for-word while someone else reads aloud, and the like. These extreme measures are useful because proofreading requires a somewhat unnatural attention to surface form and good readers are often the worst proofreaders because they supply much information that is not in fact in the surface forms at all (cf. Goodman, 1967; Goodman and Goodman, 1977; Goodman, Goodman, and Flores, 1979; Smith, 1975, 1978, 1982, 1984, 1989).

Another procedure that is integrative but fails the time requirement is the sort of multiple-choice cloze test where a list of many (say, 50 or more) words are given and must be reinserted, one by one, into a text with blanks. This task is highly integrative but may involve looking back and forth between the list and the text, and a constant rereading of the list. It may be more like solving a cross-word puzzle than normal discourse processing. Because of the frequent interruptions, in looking back and forth between text and list, and the time lapses while reading the list, it is doubtful that such a task constitutes a pragmatically viable procedure. At any rate, as the list of possible words becomes longer and longer, it is clear that the task resembles less and less the normal processing of discourse.

What was more important about pragmatic tests, and what is yet to be appreciated fully by theoreticians and practitioners is that all of the goals of discrete-point items, e.g., diagnosis, focus, isolation, etc., could be better achieved in the full rich context of one or more pragmatic tests. As a result, it was argued that the valid objectives of discrete-point theory could be completely incorporated within a pragmatic framework. However, the goal of separating each and every element of structure or skill from the whole fabric of experience was abandoned. As an analytic method of linguistic analysis, the discrete-point approach may have had some validity, but as a practical method for assessing language abilities, it was misguided, counterproductive, and logically impossible to achieve.

Another outcome of the discrete-point/integrative controversy, and the empirical research which it spawned, was a reconsideration of the almost forgotten *g*-factor of Charles Spearman (1904, 1927). This development had two sides: one statistical and the other theoretical. The statistical side of the argument was soon resolved against any all inclusive *g*-factor, but the theoretical argument has yet to be adequately considered.

Charles Spearman had observed that most intelligence tests, in his day (and it may be noted that things have changed little since then; cf. Jensen, 1969, 1980) were strongly correlated. By inventing factor analysis, then a new statistical technique, Spearman showed that it was possible to identify a single general factor underlying most IQ tests and accounting for a huge chunk of variance in all of them. The same argument could still be extended to almost all achievement, competency, and proficiency tests used in education today (see Oller and Perkins, 1978, Gunnarsson, 1978, and Stump, 1978, and for counterpoint and response, Carroll, 1983b, and Oller, 1983a; but see Gardner and Hatch, 1989 who claim to be able to measure separate "intelligences" independently). This general factor came to be known as "*g*" or "the *g*-factor". Subsequently, L.L. Thurstone (1924, 1938, 1947; also Thurstone and Thurstone, 1941) and others, argued in favor of a plurality of primary mental abilities instead of a single *g*-factor of intelligence. They never settled how many primary factors there were or just how to define them. They vacillated in the end between six and eight distinct primary factors. In more recent years Guilford's "structure of intellect" model has multiplied these factors to 120 (Guilford, 1967). More recently still, Gardner (1983, 1989, 1990), Gardner and Hatch (1989), and Walters and Gardner (1985, 1986a, 1986b) have picked up the cudgel again on behalf of multiple intelligences. While Gardner and colleagues differ in their particular list of "intelligences" from the "primary factors" proposed much earlier by the Thurstone's, there is a fundamental resemblance in both the arguments and applications of the ideas favoring profiles that look at the broad spectrum of a person's abilities rather than a single IQ score.

However, long before Howard Gardner and colleagues came to the fray, it was generally admitted (by L.L. Thurstone himself, and more recently by his student J.B. Carroll and others) that underlying any set of primary factors or secondary or tertiary ones there will still be a general factor. A recent study of language proficiency by Fouly, Bachman, and Cziko (1990) concludes that a second order general factor and a model that allows differentiated components at the first order level are both fairly good at predicting observed relations between different language measures for 334 ESL students at the University of Illinois. They refer to Carroll (1983a) who summed up both his results and those of Fouly, et al. (1990) in terms of the long term controversy over general versus specific factors in language testing research:

> With respect to whether the results support a "unitary language ability hypothesis" or a "divisible competence hypothesis," I have always assumed that the answer is somewhere in between. That is, I have assumed there is a "general language ability" but, at the same time, that language skills have some tendency to be developed and specialized to different degrees, or at different rates

> so that different language skills can be separately recognized and measured (p. 82).

Fouly, et al. go on to say, "the present study provided support for the differentiated skills hypothesis recurrent in the works of Bachman and Palmer (1983), Carroll (1983a), Farhady (1983c), and Upshur and Homburg (1983)....Similarly, the findings of this study support the claim that, in addition to differentiated language skills, there exists a general factor" (p. 16). In support of the latter model they might have cited Oller and Perkins (1978, 1980) and Oller (1983a). A general factor of language proficiency (or what has been called "intelligence," in the case of tests of primary language abilities), cannot be denied on statistical grounds (Carroll, 1983a, 1983b).

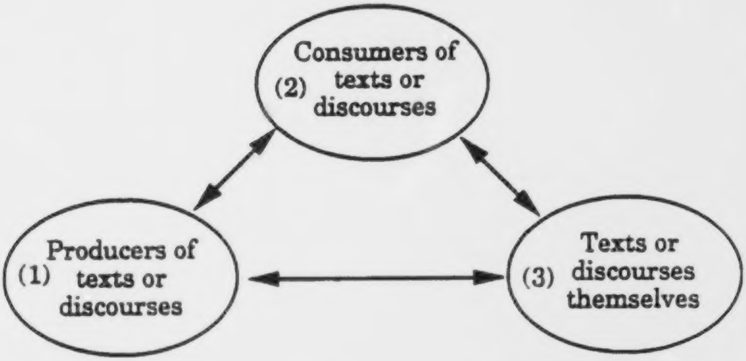
While at first multiple factors as contrasted with a general factor were thought of as mutually exclusive, this was never correct. The general factor, whimsically referred to as the Godzilla factor by Purcell (1983) could be useful in spite of the fact that it did not exhaust all of the reliable variance in a number of language tests and even though could be transformed in a variety of ways into a multitude of component factors (see Farhady, 1983c; Upshur and Homburg, 1983; Bachman and Palmer, 1983; Vollmer and Sang, 1983). Godzilla, therefore, was prematurely proclaimed to be dead (by Purcell, Farhady, and others), and certain persons set out to bury him (Alderson and Hughes, 1981; Palmer, Groot, and Trosper, 1981; Porter, 1983; Spolsky, 1983; Alderson, 1983; Hughes and Porter, 1983; Davies, 1984). But Godzilla refused to be buried. It was true that he was not quite tall and strong enough to embrace the whole world (i.e., explain all of the variance in all tests), but he was plenty large and strong enough to resist burial (Bachman and Palmer, 1983; Carroll, 1983a; Bachman, 1990; Fouly, Bachman, Cziko, 1990; Oltman, Stricker, and Barrows, 1990).

Although some researchers continue to pursue the elusive goal of resolving the general factor into its "proper" components (Sang, Schmitz, Vollmer, Baumert, and Roeder, 1986; Bachman and Clark, 1987; Bachman, 1990; Fouly, Bachman, Cziko, 1990), it would seem that a definitive division of language proficiency into its contributing components may be unachievable in principle by virtue of the fact that the multi-faceted semiotic hierarchy can be viewed from many complementary angles that logically should prove to be about equally correct (witness the findings of Fouly, et al. 1990). At any rate, the most important side of the argument is not statistical, but theoretical -- the fundamental problem is to find a coherent theory and it is certain that this cannot be achieved by purely statistical methods (see Bachman, 1990, pp. 296-358; Cummins, 1981; Krashen, 1981, 1982, 1985; Carroll, 1983a, 1983b; Upshur and Homburg, 1983). Upshur (1979), Carroll, and others have shown that the componential resolution of a general factor into a plurality of contributing components is not at all incompatible with the notion that language proficiency may be a fairly coherent and integrated totality. If we consider the meaning of total scores on tests with diverse subtests, or if we consider the fact that communicative abilities interact in complex ways to produce composite results, it is clear that both general and specific factors *must* be present in language proficiency. We will examine a few possibilities in section 3 below in this paper.

Aside from exploratory and confirmatory factoring of the traits (or theoretical constructs) that we may posit as aspects of human mental abilities or language skills (which I do not take to be the same thing, contrary to Boyle, 1987) and methods associated with particular tests, a number of interesting research reports using item response theory (IRT; following Rasch, 1980; see Davidson, 1988; Lynch, Davidson, and Henning, 1988; and Kunnan, 1990;) or multidimensional scaling (Oltman, Stricker, and Barrows, 1990; and Oltman and Stricker, 1990 following Guttman, 1965) have appeared. The common purpose of much of the research has been to sort out distinct sources of variance in language test scores. Among the widely recognized possibilities are three major sources as shown in Figure 1 below: (1) producers of discourse or text themselves differ in language abilities (and other mental abilities as well), as do (2) consumers, and as do (3) the texts or discourses (items in the case of many tests) that are both produced and understood. These three sources of variance can, of course, be further parsed up in a great variety of ways. One of the interesting and instructive avenues of research has been item response theory (IRT). Citing a single study will show how IRT can be applied to turn up unexpected sources of test item biases.

Kunnan (1990) demonstrated with an IRT approach (using a one parameter Rasch model with approximately 844 subjects) that subjects of different native language backgrounds and gender differ in performance on certain language test items depending in part on the instruction they have received probably in their major fields of study. At any rate, *differential item functioning* (DIF) was observed on the 150-item ESL Placement Examination at UCLA used in the Fall of 1987 on about 15 percent of the items. Apparently, Davidson (1988; see footnote 1 on p. 742 of Kunnan, 1990) had already shown that the test items in question met the requirement of unidimensionality in order for one parameter IRT to be applied. Based on that assumption, Kunnan found that certain grammar items focussing on the definite article, one or more prepositions, and verb tense were easier for Chinese and Japanese subjects (than for Spanish or Korean subjects), though different items (three in each case) performed differentially for the two groups. Also four vocabulary items proved significantly easier for Spanish speakers: *hypothetical, implication, elaborate,* and *alcoholics*.

Figure 1
The Three Main Sources of Variance
in Language Test Scores


(2) Consumers of texts or discourses
(1) Producers of texts or discourses
(3) Texts or discourses themselves

Since these words have Latin bases and cognates in Spanish with similar meanings, Kunnan credited native language background itself with the observed DIF for these items. Additional differences were observed for gender on 20 items some of which seemed to differ according to the major field of candidates. Items oriented toward the sciences seemed to favor males. Three items that favored females could not be accounted for. The results are interesting insofar as they show that items may be unintentionally biased against or in favor of certain groups. However, remedies for preventing this sort of bias are not clear: Kunnan, for instance, recommends that "a broad range of test content and formats" may help to reduce instructional bias. As for gender and native language biases, these are more difficult to deal with. They can be spotted on a post hoc basis with IRT, and the items can then be rewritten, but it is not entirely obvious how the author's recommendation that demographic data be elicited in advance might be used in test preparation. Certainly for items that remain unexplained even after the post hoc IRT, a demographic questionnaire or any sort of pre-screening even by members of the targeted examinees would seem unlikely to avoid the, for the moment, unexplained DIFs. The research is, in my view, nonetheless important as demonstrating the subtle kinds of test biases that can arise and the widely different sources variance that may constitute such biases.

Similar, though somewhat more specific biases for Japanese learners of English as a foreign language are demonstrated experimentally by Chihara, Sakurai, and Oller (1989). Our work used a more traditional repeated-measures approach but predicted in advance what sorts of items in a cloze passage were biased against Japanese learners of EFL. Because Japanese subjects were compared against themselves in a repeated measures design, the variance of interest in particular items can be attributed specifically to the cultural or experiential background of the subjects tested. Two cloze passages were each presented in two forms: each passage appeared in an unmodified (biased) form and in a modified (reduced bias form). The method of modification was to change unfamiliar place names in the U.S. and Greece to familiar ones in Japan, and one instance of a mother kissing her son was changed to hugging (which is acceptable in Japanese culture). The results showed a significant advantage overall favoring the modified texts in spite of the fact that all else was left unchanged. The results, though based on an entirely different experimental procedure, agree with those of Kunnan (1990) using IRT, in showing that items may function differentially according to the background of subjects.

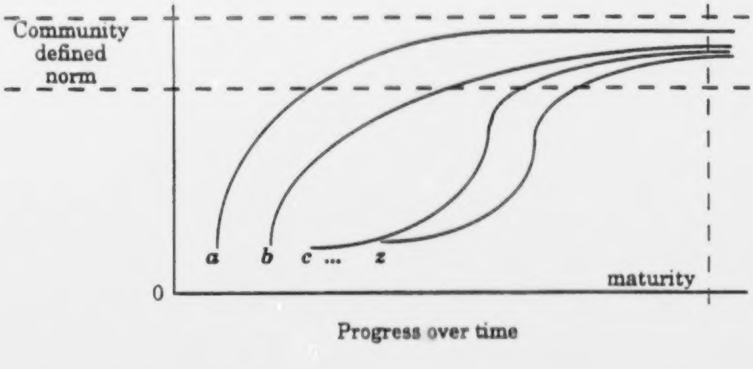
A rather different application of IRT comes from Lynch, Davidson, and Henning (1988). While Kunnan (1990) was interested in variance across items, Lynch, et al., focussed on variance within persons (on a different form of the same UCLA ESLPE examined by Kunnan). Lynch, et al., wanted to determine if variance within persons could also be regarded as unidimensional. It had been determined in several prior studies that variance across items tended to be unidimensional. Both person variance and item variance need to be unidimensional in order for one-parameter Rasch models to be optimally applicable. Like Oltman, Stricker, and Barrows (1990) -- who used a different approach, multidimensional scaling (following Guttman, 1965) -- the evidence obtained by Lynch, Davidson, and Henning (1988) seemed to show that unidimensionality may not be achieved until language learners gain some maturity in the target language. Their conclusion expresses this idea negatively: with reference to violations of unidimensionality, they say that their results seem to support the notion that such violations are more serious at the lower end of the ability continuum (p. 218).

Citing Oltman and Stricker, Lynch, et al. note that the few dimensions detected tend to merge into a larger primary dimension at the upper end of the ability scale (p. 207).

This same observation has been made by Oltman, Stricker, and Barrows (1990) on the basis of a different statistical technique (multidimensional scaling).

Whereas Lynch, et al., studied responses of 678 subjects taking the UCLA ESLPE in the Fall of 1987, Oltman and colleagues studied 53,169 subjects who took the *Test of English as a Foreign Language* in May of 1985. These results give fairly persuasive evidence that whatever factors or dimensions language proficiency may resolve into probably do vary dynamically over time just as Clifford (1980) and Lowe (1980) predicted they would. In fact, Figure 2 suggests an abstract idea of the sort of thing that appears to be happening with the TOEFL and with the UCLA ESLPE as well. Whereas in the early stages of second language learning, distinct dimensions of listening, writing, and reading ability may be observed (and these may even resolve into further sub-component traits or categories), as learners progress to a more mature, native-like capacity in the target language, it seems that the diverse dimensions (factors, traits, or whatever they may be called) tend to converge to a more unidimensional structure.

Figure 2
Hypothetical Convergence of Arbitrarily Designated Factors or Dimensions Designated a, b, c, ...z (traits, methods, or whatever) of Language Proficiency Viewed Over Time until Maturity is Attained.



A tentative hypothesis may be offered: Perhaps the various dimensions (whether attributed to persons or to items) that are sorted out by language tests (and observed in some detail through multidimensional scaling techniques) tend to converge on some more or less well-determined norm that is defined by the community of users who know and use the target language in question for the sorts of purposes that the language tests inadvertently characterize. There are good theoretical reasons to suppose that some sort of normative convergence must in fact occur in "normal" language acquisition. Whereas learners may vary considerably in the rate and degree of initial success in mastering all of the diverse aspects of a language system, the sounds and meanings of words, the syntax and semantic values of phrases and clauses, not to mention pragmatic applications in experience, must all *tend* toward more or less standardized norms in order for communication to be possible across the diverse members of any given language community. It is precisely in this sense, I believe, that language tests must always to some degree be normative in principle. Criterion-referencing is not ruled out, but it will necessarily be incomplete unless supplemented by norm-referencing (i.e., specifically to the norms of the language community in question). Languages, whatever else they may be, are intrinsically, norms of symbolic behavior. We will return to this idea in section 3 below, but first it may be useful to examine some of the broader research on the measurement of human abilities in order to appreciate better the special role played by language abilities.

(2) Review of Educational Measurement

Modern variants, of the analytic approach typified by the discrete-point foreign language testing of the 1960s can still be found in abundance in the general literature of educational measurement. Kagan (1990) complains about the "atomistic view of effective teaching that emerged from the process-product research of the 1970s" as well as the mistaken notion that a teacher's competency can be defined entirely in terms of a "laundry list of behavioral objectives" (Howey and Zimpher, 1989; Kagan, 1990, p. 419). Of course, a review of the literature shows that the laundry-lists have not been limited to behavioral objectives for teachers but have been extended to every domain of the curriculum and every sort of testing -- including tests aimed at intelligences, achievement, bilingualism, language disorders, etc.

Nowhere is the atomistic, discrete-point approach more apparent than in the literature about how to construct "items." In fact, the analytic, taxonomical philosophy (reflecting little influence as yet from the Chomskyan revolution; e.g., see the numerous references to the taxonomy of Benjamin S. Bloom still prevalent in the literature) continues to hold sway in most educational and psychological testing. For example, Roid and Haladyna (1982) describe "the heart of what is currently known as CR [criterion-referenced] testing" as the notion that "a domain-based interpretation is possible only when a domain or universe of items has been created and the test is based on a sample from this domain" (p. 28). A domain, according to such thinking, is conceived of as a list of potential items from which a sample is drawn in constructing a test. Roid and Haladyna (1982) attribute to Bormuth (1970) the idea that a technology of item writing might "be based on the transformation of sentences into questions" (p. 99). A domain, by this view, is a list of sentences. They acknowledge that the whole idea of sampling from a domain of sentences is susceptible to "serious objections" that arise in connection with "the meaningfulness of definable universes" (p. 34).

There are really two problems here: modern linguistic theory shows that the number of sentences in any given domain of interest for practical purposes is non-finite, and it also shows that any known method of algorithmically generating sentences will produce a great deal of nonsense. Roid and Haladyna (1982), without apparently understanding the linguistic necessities, say "there is a chance for endless mapping sentences, facts, and facet elements, with lack of agreement among developers being a major detriment to progress" (p. 132). The non-finiteness of sentences about any given subject matter renders the idea of a "randomly selected representative sample" uninterpretable, and the abundance of nonsense that would be generated by any known algorithmic procedure makes that approach relatively unappealing. Further, the recommendation (of Bormuth, 1970, cf. Roid and Haladyna, 1982, p. 92) that all possible items in a domain be specified is logically (in principle) unattainable. For these and other reasons, I still believe (cf. Oller, 1979, pp. 32-33) we need to look for an approach to educational and psychological testing that assesses the relative efficiency of a generative system (i.e., the symbolic system itself) rather than attempting to representatively sample from an unattainable listing of an infinitude of demonstrably infinite universes of particular sentences or test items. When the focus is shifted from a list of items (a poor characterization in any case of any non-finite domain of sentences) to the generative basis which underlies the representations that constitute that domain, we have some hope of achieving both reliability and validity. While approaches to educational and psychological measurement have yet to appreciate the purely theoretical implications of the Chomskyan revolution, happily a movement toward more pragmatic, holistic, testing is nonetheless discernible.

Whereas Roid and Haladyna (1982) view individual test items as the "basic building blocks of tests" (p. ix), they implicitly take into account the contrast between (1) discrete-point theory where individual items are matched with some abstract trait and a more pragmatic approach where (2) the tester/teacher thinks in terms of "a theory of the relations between a test and other variables in the real world (a nomological network)" (p. 8). The latter approach would seem to address the fundamental problem of pragmatic mapping (also known as abductive reasoning) to which we return in part 3 below. It is also refreshing to read in Roid and Haladyna (1982) that "testing is viewed as a part of instruction and not a separate operation" (p. 30). In this they follow the lead of people like Eva L. Baker (1980) who argues for a comprehensive "integrating" model of "teaching-learning-assessment" (p. 14) where the various activities are merely viewed from different perspectives, but not as distinct and separate entities apart from the whole context of education. It is the articulation of a theoretical basis for such holistic, nomological, or pragmatic approaches, the author will argue in section 3, that is most needed.

The author agrees with Gardner (1990) who cites Chomsky (1975) in support of the idea that the acquisition of various representational abilities -- though not always the more abstract academic ones that Gardner calls "literacy, numeracy, and critical thinking" -- is natural and normally proceeds without a hitch. "Given environments that are not grossly impoverished, all children will learn how to speak and understand their native languages (and other languages in their surround) with ease and facility; acquire basic understandings of the operation of the physical world (the constancy of matter, the principles of cause and effect); understand key aspects of the social world (the way to convince another individual, the detection of benevolent or malevolent motivation); and use a range of symbolic codes, such as those involved in picturing, gesturing, and making music, in order to express and derive meanings" (pp. 89-90). Following Chomsky, Gardner acknowledges that not only do children normally accomplish such things without special tutelage, but that "adults do not know *how to teach* [his italics] many of the most important forms of knowledge which every normal child acquires" (p. 90.

Gardner in all of his recent writings stresses the partial independence of "intelligences." He says, "While such areas as reading, or studying history, or composing music may well be characterized by stages of competence, the stages found in one domain may have little resemblance to, or correlation with, those regnant in other domains... even in those areas of learning which appear to be universal, all forms of learning do not develop in synchrony. Rather, human beings differ in the manner in which, and the speed with which, they express various mental capacities or 'intelligences' " (pp. 90-91). He points out that learners often exhibit what may be called "U-shaped" growth or learning curves. They seem to acquire a concept but fail to generalize it appropriately to new contexts or over-generalize it to contexts where it does not work. He argues that what is missing in such cases is what he calls "connecting tissue" that would relate abstract symbolic representations to the world of experience more articulately and more completely. In my terms, what is missing is the sort of pragmatic mapping that all genuine learning requires. Too much discrete-point, surface oriented materials passes for curriculum and yet does not achieve much effect. Students remain without the pragmatic linkages to their experience that would make sense of such materials.

Gardner (1990) says that "so long as testing is geared exclusively to 'school knowledge' " -- i.e., the surface-oriented, discrete-point, unintegrated variety -- the "credentials provided by the school may bear little relevance to the demands made by the outside community" (p. 93). To remedy the situation, he is concentrating his efforts on developing "new forms of assessment which are sensitive to particular intelligences and which can document the kinds of learning that take place 'in context' in which students carry out projects of some scope" (p. 104; also see Gardner, 1989; and Gardner and Hatch, 1989). He says that "finding the topic or skill with which one feels 'connected' is the single most important educational event in a student's life" (p. 104; also Gardner and Walters, 1986a).

In coming to his eventual list of seven basic intelligences, Gardner and colleagues examined several sources in the literature: (1) normals (2) pathological and special populations including such cases as autism, savantism, and learning disabilities. Gardner and Hatch (1989) claim that it is possible to escape the biased confines of "linguistic and logical skills" by developing what they call "intelligence fair measures" that "seek to respect the different modes of thinking and performance that distinguish each intelligence. Although spatial problems can be approached to some degree through linguistic media (like verbal directions or word problems), intelligence-fair methods place a premium on the abilities to perceive and manipulate visual-spatial information in a direct manner. For example, the spatial intelligence of children can be assessed through a mechanical activity in which they are asked to take apart and reassemble a meat grinder....Although linguistically inclined children may produce a running report about the actions they are taking, little verbal skill is necessary (or helpful) for successful performance on such a task" (p. 6). Here Gardner and colleagues seem unaware of relevant research by A.R. Luria (1959, 1961, 1979; also Luria and Yudovich, 1959). Luria showed that the integration of verbal skills with certain motor tasks was essential to successful performance of those tasks for children at an early stage of development (e.g., being able to push a button consistently when a green light was on but not when a red light was on).

Serendipitously, in keeping with caveats of pragmatic testing, however, Gardner and colleagues (e.g., Gardner and Hatch, 1989) recommend holistic, highly pragmatic assessment procedures: "even at the preschool level, language capacity is not assessed in terms of vocabulary, definitions, or similarities, but rather as manifest in story telling (the novelist) and reporting (the journalist). Instead of attempting to assess spatial skills in isolation, we observe children as they are drawing (the artist) or taking apart and putting together objects (the mechanic)" (p. 6). Their approach they admit "blurs the distinctions between curriculum and assessment" (p. 5) but this surely we must applaud. It falls in line with recommendations coming from a number of quarters these days for blurring not only the lines between teaching and testing but also between the school, home, and community (Simich-Dudgeon, 1987; and Quintero and Huerta-Macías, 1990).

Parent involvement is stressed by Quintero and Huerta-Macías (1990): they say, "the positive impact of parents' involvement in their children's education is well documented (here they cite among others Simich-Dudgeon, 1987 and Wells, 1986)" (p. 307). They point out that "instructional activities must not only be interactive in nature, but also rich in cultural meanings, comparisons, and critical analysis for making classroom and out of classroom connections" (1990, p. 312). Or, as Freire and Macedo (1987) put it, "the command of reading and writing is achieved beginning with words and themes meaningful to the common experience of those becoming literate, and not with words and themes linked only to the experience of the educator" (Quintero and Huerta-Macías, 1990, p. 42). Or, from a different angle, Smith (1989) says, "individuals become literate not from the formal instruction they receive, but from what they read and write about and who they read and write with" (p. 353). Quintero and Huerta-Macías argue for a "whole language approach" (citing among others Bruner, 1984; Goodman, 1986; and Smith, 1984) they define it: "the whole language approach to language learning emphasizes that language be taught naturally as it occurs within any social environment instead of segmenting it into bits and pieces" (1990, p. 307). They recommend an experience-based approach appealing to the rich existing experiences of the family (Auerbach, 1989).

However, it is important to keep in mind, as Miller (1990) stresses that the broader and deeper view of literacy that whole-language approaches advocate also suggests connections that have too long been neglected: "Literacy viewed from the perspective of communication arising from shared activities with meaningful others cannot be separated from the issues of intelligence, learning, and language...literacy becomes entwined with how and what people know -- with intelligence" (p. 2). When this broader view is assumed, we may hope for better results in education. Quintero and Huerta-Macías (1990) conclude: "In sum, because Project FIEL [Family Initiative for English Literacy] stresses language use in meaningful context, the student's needs, wishes, and past experiences naturally become the teaching methodology, and flexibility of the curriculum is a natural result. Program goals are reached by students, parents, and teachers working together through interaction and learning for real-life needs. Finally, the experience of the project indicates that when social context is attended to in a positive way and the dignity of the learner is upheld, learning occurs" (p. 312).

By using context-rich materials and activities that engage children more fully and challenge their "intelligences" more specifically, Gardner and Hatch (1989) report higher motivation and evidence of a greater diversity of abilities. They report on a study in 1988-1989 with 20 preschool children who were tested on "story telling, drawing, singing, music perception, creative movement, social analysis, hypothesis testing, assembly, calculation and counting, and number notational logic" (p. 8). The authors conclude that only the activities requiring "logical-mathematical intelligence" proved significantly correlated with each other (r = .78, p < .01)." Their analysis, however, may be more detailed than the small number of preschool subjects in their study would justify. In a follow-up with first graders, 15 in all, again the conclusions are perhaps too general to be sustained by the small number of observations involved, but some evidence is provided showing that children do differ in expected ways on the different intelligences posited.

Walters and Gardner (1985) say that "each intelligence" (of the seven Gardner had previously identified) "must have an identifiable core operation or set of operations": for example "one core of Linguistic Intelligence is the sensitivity to phonological features" (p. 4). They say,"While it may well be possible for an Intelligence to proceed without an accompanying symbol system, a primary characteristic of human intelligence may well be its gravitation toward such an embodiment" (p. 5). Of course, if we follow C.S. Pierce, we must suppose that a sign system of some sort is prerequisite to any intelligence whatever. Here is where some additional theoretical development, I believe, is needed.

Another trend in the general educational-psychology literature that corresponds to a move away from atomistic analytic approaches and toward more holistic pragmatic procedures can be seen in studies of language disorders and learning disabilities. Audet and Hummel (1990), for instance, give an interestingly pragmatic analysis of the discourse of a nine-year-old boy diagnosed as language-learning disabled and behaviorally disordered. In general, they followed the discourse analysis procedures recommended by Damico (1980, 1985a, 1985b, and 1991). Although, Adams and Bishop (1990) and Bishops and Adams (1990) did a less fine-grained analysis (see their comparison of their own with Damico's approach on p. 260), like Damico (1985b) they were also able to show substantial reliability for judgments of pragmatic appropriateness. The shared point in all these cases, however, was to give greater attention to pragmatic aspects of discourse (an approach also advocated by Miller, 1990 and by Prutting and Kirchner, 1987).

(3) Language Proficiency in Relation to a Theory of Intelligence

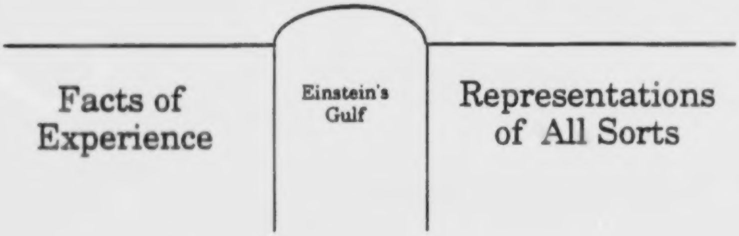
The bulk of the research on intelligence measurement per se is only tangentially relevant to a theory of language proficiency in relation to a comprehensive model of intellect. The IQ measurement research has been limited by its taxonomic character from the beginning and has scarcely begun to consider the full implications of the Chomskyan revolution. The fact is that psychology and psychometrics are yet to feel the force of generative theory. Taxonomic models, e.g., Guilford's "theory of intellect" (1967) and Bloom's taxonomy (1976; also Bloom and Krathwohl, 1977), are not merely out of date, they are either incorrect in fundamental ways, or else, the generative conception of grammar is entirely misguided. At any rate, the taxonomies, when compared against generative theories, cannot compete in scope or power. They are logically too impoverished to even begin to account for the facts of human language ability not to mention other semiotic capacities.

On the other hand, the generative conception of grammar was implicit in much work before the Chomskyan era. Such a conception was apparent in Saussure's advocacy of a general theory of "semiology." Before that, C. S. Pierce [1839-1914], a scientist characterized by Ernest Nagel in 1959 as "the most original, comprehensive, and versatile philosophical mind this country has yet produced," had written the equivalent of 104 volumes of 500 pages each in octavo, focussed primarily on the theory of semiotics. Pierce, more than any other scholar, worked toward a general theory of representations. The essence of Pierce's conception of the relation between language and intellect is suggested by Albert Einstein (1941):

> Everything depends on the degree to which words and word-combinations correspond to the world of impression.
>
> What is it that brings about such an intimate connection between language and thinking? Is there no thinking without the use of language, namely in concepts and concept-combinations for which words need not necessarily come to mind? Has not everyone of us struggled for words although the connection between "things" was already clear?
>
> We might be inclined to attribute to the act of thinking complete independence from language if the individual formed or were able to form his concepts without the verbal guidance of his environment. Yet most likely the mental shape of an individual growing up under such conditions would be very poor. Thus we may conclude that the mental development of the individual and his way of forming concepts depend to a high degree upon language (1941, in Oller 1989, p. 62).

Pierce and Saussure, presumably for similar reasons, agreed in this assessment. Both of them contended that language is the canonical semiotic medium and that by the systematic study of it we should be able to optimize our understanding of representational ("semeiotic," Pierce's term, or "semiological," Saussure's term) processes in general. More recently Noam Chomsky has urged the same program. He wrote in 1972: "One would expect that human language should directly reflect the characteristics of human intellectual capacities" (p. ix).

Figure 3
Pragmatic Mapping of Representations onto the Facts of Experience via Abductive Reasoning



Figures 3-7 elaborate on this central theme. Figure 3 pictures the primary representational problem as outlined in the above remarks by Einstein, and more fully by Pierce in the Nineteenth Century. On the left hand side of the diagram the raw uninterpreted facts of experience are pictured; on the right hand side, representations of them. The question for a theory of intellect is how the connection between the two realms is accomplished. This in a nutshell is the pragmatic mapping problem, or in Pierce's words it is the problem of abductive reasoning. It is construed, in the theory under consideration, to be the primary problem of intelligence.

Einstein described this problem and defined the "gulf" as shown in the following lines:

> ...the concepts which arise in our thought and in our linguistic expressions are all -- when viewed logically -- the free creations of thought which cannot inductively be gained from sense experiences. This is not so easily noticed only because we have the habit of combining certain concepts and conceptual relations (propositions) so definitely with certain sense experiences that we do not become conscious of the gulf -- logically unbridgeable -- which separates the world of sensory experiences from the world of concepts and propositions (1944, in Oller 1989, p. 25).

Readers familiar with Chomsky's work will not fail to see the profound similarity between what Einstein says here and what Chomsky has said many times elsewhere. The idea that true representations are validly connected with whatever they purport to represent, otherwise known as the correspondence theory of truth, is foundational to what Einstein is saying in the immediately preceding quotation. Moreover, it is implicit in many of the remarks of educators concerning the need to relate what is talked about in the classroom to the actual, real-life, real-world experience of students both in and out of the classroom.

Probably the main reason that the Peircean or Einsteinian view of reality has not been more widely accepted by scholars is owing to a peculiar skepticism about our knowledge of the external world that still prevails in much modern thinking and education. MacNamara (1989) shows that modern approaches to human representations often assume an extreme variety of such skepticism. In reviewing a collection of works representing some of the most widely read theoreticians of the present decade (Umberto Eco, Roger Schank, Ray Jackendoff, George Lakoff, and others), MacNamara (1989) complains that "the collection radiates skepticism about the capacity of the mind to know reality" (p. 350). While some of the authors see mental models as mediating between representations and the external world, others see them as being only in contact with themselves. Now it follows that if mental representations have only themselves or other mental representations as their ultimate objects, thinking is quite independent of any external reality, and must be regarded as essentially unrelated to our actions. Common sense and all logic rejects this extreme view. On the contrary, we suppose that people are responsible for their actions in a way that inert objects and unreasoning organisms are not and that the responsibility is based in the linking of representations with corresponding facts that have an independent reality of their own.

When a representation corresponds faithfully to a fact we say that the representation is true of that fact. This is the layman's definition of truth and it does not differ in any essential respect from that of the scientist. However, some skeptics suggest that the very correspondence of a representation with a factual state of affairs is itself a fiction. For instance, Umberto Eco capsulizes this view in his chapter title, "On truth, a fiction" (in Eco, Santambrogio, and Viola, 1988). While C.S. Pierce, whom Eco claims to follow, saw truth as a purely abstract quality of representations (which would give it the same immaterial quality as any fiction -- thus making it fiction*al*), Pierce did not assign any extra degree of reality to material entities so the abstractness of truth would not detract in the least from its reality. On the contrary, while physical things, owing to the laws of thermodynamics come into existence in space and time, grow old, wear out, and are no more, the truth of any representation (e.g., that these words were written by yours truly in Albuquerque, New Mexico, at about 2:25 in the afternoon on August 4, 1991) is an eternal fact. It does not change over time. Therefore, for Pierce, truth was not a fiction, though it has the same abstract *quality* as a fiction. The difference between these views is like that between a libertarian skepticism on the one hand, and a responsible pragmatism (or what Pierce called "pragmaticism" to distinguish his views from those of William James and John Dewey) on the other.

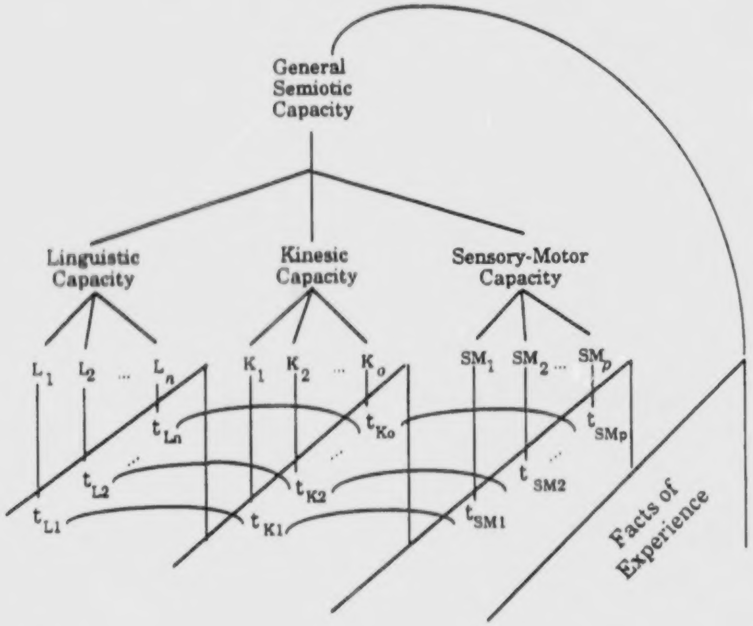
I have mentioned skepticism because it is probably the prevailing view among theoreticians of the twentieth century in spite of the fact that the typical school teacher takes a more realistic approach. For instance, when educators and parents speak of relating classroom activities to the real world, they presuppose that a real world exists and that we have some more or less valid knowledge of it. Therefore, if whole-language, experience-based, socially relevant curricula are actually possible, the extreme variety of skepticism must be wrong.

Figure 4 elaborates on the model by proposing a hierarchy of three distinct kinds of representational capacities: linguistic, kinesic, and sensory-motor. According to Pierce, the language capacity is fully abstract and may be used to represent any imaginable, or even unimaginable idea whatever. We may at least speak of the unimaginably fantastic. The kinesic, gestural, sort of representation is intermediate. It is conventional and arbitrary to some extent, but may also involve iconic (analogical) elements. For instance, a brandished fist suggests more or less iconically the act of punching someone, but it may by convention acquire a rather different meaning -- e.g., it may be a sign of solidarity or brotherhood.

Or consider the fact that Americans and most western Europeans indicate themselves kinesically by pointing roughly at their own sternum (the center of the chest) with the right index finger or thumb of the right hand. Japanese, however, point to themselves by touching or pointing toward their nose with the right index finger, palm turned inward toward the body. Each of these gestures has its conventional aspects as well as its universal basis in the ego-reference point. The latter is not a mere convention since it is physiologically impossible for a perceiver to have any other primary reference point. (Without the notion of one's own self, it would be impossible to credit any other self with existence or to differentiate the self from any other person; see Pierce, in Moore, et al., 1984, pp. 201ff.)

Sensory-motor representations on the other hand are more or less directly, and iconically, related to the facts of experience. Persons skiing down a mountain not only represent the terrain ahead in a continuous flow of images but must also represent at some level body postures and internal commands for motor adjustments in order to control body and skis to accommodate the slope beneath them.

Figure 4
The Semiotic Hierarchy in Terms of
Principal Systems of Representation


General Semiotic Capacity
Linguistic Capacity
Kinesic Capacity
Sensory-Motor Capacity
L_1 L_2 ... L_n
K_1 K_2 ... K_o
SM_1 SM_2 ... SM_p
t_{Ln}
t_{Ko}
t_{SMp}
t_{L2}
t_{K2}
t_{SM2}
t_{L1}
t_{K1}
t_{SM1}
Facts of Experience

As Pierce showed with unassailable logic and meticulous phenomenological analysis, sensory-motor representations are analogues, copies, or icons of the facts they represent and, as such they are degenerate. If we look away from an object, its image quickly fades. Details are lost or may be wrongly reconstructed in the mental picture.

Kinesic representations are similar in character, yet may contain an added conventional element. For example, Europeans and westerners in general are apt to point with the index finger to call attention to an object or event. Navajos achieve the same purpose by extending the lower lip. Indexes, a second kind of representational form, are reactionally degenerate. They are not generally so explicit as to rule out the possibility of our noticing the wrong thing which amounts to failing to notice whatever was pointed at. Pierce called this special kind of degeneracy, reactional and distinguished it from the qualitative degeneracy of icons.

BEST COPY AVAILABLE

Linguistic representations by contrast achieve a higher level of abstraction and a closer approximation to validity. It is true that they must involve icons and indexes to the extent that they are synthetic in character, i.e., to the extent that they inform us about actual experience, but their fundamental character pertains to their abstractness and near independence of anything external to them. While linguistic forms that depend on sensory-motor representations of non-linguistic states of affairs (e.g., factual or fictional contexts), or that appeal to indexical or deictic relations (e.g., pointing or naming or referring) involve the same kinds of degeneracy associated with icons and indexes respectively, the purely semantic values associated with words and propositions are quite impervious to either of those sorts of degeneracy. For instance, our concept of mortality does not deteriorate from one moment to the next in the way that our recollection of a scene does. That is, the semantic value of a word or proposition is not qualitatively degenerate. Nor does our idea of mortality depend on any particular instance of it that might be singled out for attention (e.g., the fact that Socrates died). In fact our abstract concepts (or the abstract meanings of words, propositions, and texts/discourses are not at all reactionally degenerate in the way indexes are. Therefore, Pierce argued, symbols are relatively genuine, i.e., pure and valid by comparison to icons and indexes.

In addition to the fact that linguistic representations are primarily symbolic while gestures have an intrinsic indexical quality in many instances and sensory-motor representations are largely iconic, a few more words need to be said about the three main categories of semiotic systems. Because of their greater abstractness and symbolic character, linguistic representations and their underlying forms embody certain cognitive powers of reasoning that the other two major classes of representations are not capable of achieving. For instance, there is no way that any iconic representation can express adequately the notion that human beings are mortal. Nor is it possible to express that idea strictly speaking in an index or any other sort of mere gesture. An abstract grammatical system capable of expressing a practical infinity of subject-predicate relations, negations, conjunctions of ideas, and the like is required to express fully what is meant by the fact that human beings are mortal or any other similarly complex abstract proposition. However, kinesic and sensory-motor representations also have certain special properties. For instance, an iconic representation, such as a visual representation of a scene, cannot be quite perfectly translated into words. The Chinese aphorism that a picture is worth a thousand words is an understatement. A picture is worth many more than a thousand words. Similarly, gestural systems have unique capabilities. Just as a picture is worth a thousand words, a single look, a facial expression or tone of voice may speak volumes. Affective information, it seems, the emotive side of human experience is far more effectively conveyed in facial expression and tone of voice than it ever could be in words or images alone.

Therefore, each of the three major semiotic systems has its own special capabilities. Still, it must be said that language reigns supreme as commanding the greatest degree of independence from the material world and also, by far, the greatest degree of generality relative to its scope. We cannot visualize, hear, smell, taste, or feel everything we can talk about, nor can we express in paralinguistic mechanisms every idea we can talk about. On the other hand, we can talk about absolutely anything that is conceivable. Anything beyond our capability to represent in some oblique manner in words is simply beyond our conception altogether.

So much for the three general headings under the overall intellectual ability termed "General Semiotic Capacity" in Figure 4. It remains to explain the terms subordinate to each of these. Under "Linguistic Semiotic Capacity," an ability that is believed to be innate and species specific to human beings, come terms that correspond to the grammars of particular language systems, L_1, L_2, through L_n. These systems, to the extent they are not already specified by innate knowledge of universal grammar, must be acquired if they are to be known at all. Each in its turn corresponds then to a class of textual representations in experience, t_{L1}, t_{L2}, through t_{Ln}. These terms stand for the texts, for instance, that conform to one's primary language, or second language, and so forth. For monolinguals, there will be no L_2.

The same sort of hierarchical arrangement is hypothesized under the "Kinesic Semiotic Capacity." It too is expected to be largely innate though not entirely species specific to human beings. Again, the universal kinesic capacity dominates (or branches into) a plurality (or at least a potential plurality) of subordinate acquired systems. Each of these subordinate systems dominates a class of texts or representational forms in experience, and these tend to be loosely tied to linguistic texts. For example, English speakers are apt to accompany the statement that a certain person is about "so tall" with a corresponding gesture, palm down, hand extended. A speaker of a different language may use a quite different conventional gesture for the same purpose.

More importantly, research shows that the sequence of gestures is delicately coordinated with the sequence of linguistic forms and meanings. According to research by Condon and Ogston (1971) this is true not only of the speaker but also of the audience to such an extent that their body movements appear to be under the control of one and the same puppeteer.

The case for Sensory-Motor Capacity, if anything, is more dramatic. There is no question that much of our ability to perceive the world and our body as part of it, must be innate (cf. T. G. R. Bower, 1971, 1974; also the Chomsky and Piaget debate in Piatelli-Palmarini, 1980 and comments from the other participants). However, every normal person operates in ordinary experience by so many routines and patterns that it would be impossible to estimate how many distinct sensory-motor systems an ordinary individual possesses. There are sensory-motor programs for almost every imaginable aspect of routine experience, chewing gum, brushing your teeth, grooming in general, dressing, tying your shoes, driving a car, riding a bicycle, playing basketball, going to class, giving a talk, writing a letter, typing one, talking on the phone, etc., and each of these routines is divisible into subroutines of a great variety.

To the extent that such programs can be made explicit as rule-governed systems, they are like grammars of natural languages. They also have their own sensory-motor texts, t_{SM1}, t_{SM2}, and so forth. For instance, our ability to recognize a game of basketball and to distinguish it from a tennis match, or to distinguish either of these from a boxing match, is dependent in part on our knowledge of the corresponding sensory-motor systems. But none of these knowledge systems is the same as an actual game of basketball, or tennis, or a particular boxing match. Yet, the general rule-systems underlying the particular manifest forms (t_{SM}'s in Figure 4) are at least as distinct from each other as are the diverse "textual" manifestations. Sensory-motor texts, in their turn, are also coordinated in ordinary experience in delicately articulate ways with kinesic and linguistic texts.

Because the information processing approach to the development of semiotic systems over time is discussed in Damico and Oller (1991) along with a detailed analysis of some of the empirical evidences in favor of the theory, I will merely summarize those evidences here and will skip over much of the discussion given there (Damico and Oller, 1991) of the theory from an information processing point of view.

Empirical evidence in favor of the theory sketched out includes first, a plausible explanation of our ability to translate information from one semiotic system into another. Each of the universal systems of knowledge (and no claim is made as to the completeness of the ones postulated, only their necessity) though distinct, is related to the others through the domination of the general capacity, and each also subordinates one or more particular systems that are acquired and are to some extent conventional in character. For example, the acquisition of the primary language at once fleshes out the universal aspects of language that are realized in that system and at the same time results in the addition of conventional features that are unique to the primary language. Much the same will be true in the acquisition of the kinesic system that accompanies the first language. Our ability to translate information from one system more or less adequately into another is indicative of the underlying general capacity that connects the different quasi-independent modules or in Gardner's terms "multiple intelligences." We can talk about what we see or describe in words the meaning of a gesture, facial expression, or tone of voice. Or, we can visualize a scene as someone else describes it, imagine a facial expression, tone of voice, or the like based on a linguistic representation. Paraphrase is included as a special case of such translations. We can also paraphrase meanings that have been expressed in a certain surface form by putting them into other surface forms that give more or less the same result. For instance, the statement that "Men are mortal" may be paraphrased by saying that "All humanity must ultimately face death" or that "Mortality is a trait of human beings," etc. Translation across distinct language systems, e.g., "Los hombres son mortales" or "La mortalidad es una de las cualidades de los hombres," or translation into any language or other form that can be imagined, is ample evidence in favor of a general factor of semiotic capacity. Apart from such a general capacity, such translations (even quite imperfect ones, much less fully satisfactory ones) would be inexplicable.

I agree with Roid and Haladyna (1982) as well as Anderson (1972) who recommend the use of paraphrase in the testing of comprehension of prose materials in a school curriculum. Roid and Haladyna (1982) say that "the reason for using paraphrase [in testing] is to ensure that students have truly comprehended the ideas... that they have not just recalled the wording at a surface level" (p. 91). They quote Anderson (1972): "to answer a question based on a paraphrase, a person has to have comprehended the original sentence, since a paraphrase is related to the original sentence with respect to meaning but unrelated with respect to the shape or sound of the words" (p. 92). My point, however, is a little different than theirs as I am stressing the fact that all comprehension of a semiotic sort involves a sort of paraphrasing or translation into a different semiotic medium. This idea comes from Pierce and was viewed by Roman Jakobson (1980) as the special genius of the whole Peircean perspective on semiotics and linguistics. Jakobson commented that "the translation of a sign into another system of signs" as a definition of the process of interpretation was "one of the most felicitous, brilliant ideas which general linguistics and semiotics gained from the American thinker" (p. 35).

Now here is where the theory of Walters and Gardner runs into a difficulty: if there were really independent "intelligences," it should not be possible to translate very well from one to another. They, of course, admit that it is possible to do some such translation and yet at the same time see this as a bit of a "conundrum." They give an example of a non-mathematically inclined child who must master some mathematical principle. They say, after the mathematical approach fails, "the teacher must attempt to find an alternative route to the mathematical context -- a metaphor in another medium. Language is perhaps the most obvious alternative, but spatial modeling and even a bodily-kinesthetic metaphor may prove appropriate in some cases.

In this way, the student is given a *secondary* route to the solution... perhaps through a medium that is relatively strong for that individual” (p. 20). What this potential detour to the difficult mathematical principle shows is that it must be possible to some degree to translate between the different symbolic media. However, they surmise that “there is no *necessary* reason why a problem in one domain *must be translatable* into a metaphorical problem in another domain... as learning becomes more complex, the likelihood of a successful translation diminishes” (p. 20). They assert, “the mathematical principle cannot be translated *entirely* into words (which is a linguistic medium) or spatial models (a spatial medium)” (p. 19). However, no proof of this has been offered, and Peircean theory shows that one of the properties of truly symbolic systems is their relatively perfect intertranslatability. While we cannot translate from an icon to an index, nor vice versa, nor can we always translate from a symbol to either an icon or an index, we can always translate from one symbol to another, and there is no limit to the accuracy of such symbolic translations. Furthermore, all indexes and icons are more or less translatable into symbols, though the reverse is sometimes impossible. How, for instance, would you adequately represent the mortality of human beings by pointing to something in particular? Or what icon would show the full meaning of the symbolic proposition that humans are mortal? On the other hand, a verbal description may suggest an icon just as it may suggest a particular index. In fact, verbal descriptions can literally include icons and indexes within them so as to more or less completely usurp their special representational capacities.

The fact that fairly complex translations are meaningful is demonstrated in the sort of research exemplified by Nolen and Haladyna (1990). They focussed on two types of study strategies that encourage “deep-processing” (their term): *elaboration* (e.g., “figure out how it fits in with what you learned in class”) and *monitoring* (“asking yourself questions while you read to make sure you understand”) (p. 117). They argue that “if students think the teacher wants them to understand material and relate it to their own lives, as well as to think creatively and independently about it, they will come to value strategies (like monitoring and elaboration) that lead to those goals” (p. 119). Now if translation of the sort that takes place between distinct semiotic media were not fairly good, it is difficult to see how “deep-processing” would relate to all of the diversity of concepts, illustrations, photographs, texts, experiments, etc. that constitute the curricular bases for learning about science. In fact, the whole thesis of experience-based, socially relevant, whole language education, is grounded in the implicit assumption that meaningful connections and translations across distinct semiotic media are not only possible but more normal than the traditional analytic separation of those media into separate and independent categories.

Another evidence of the connectedness of the various disciplines summed up in Gardner's terms "literacy", "numeracy", and "critical thinking" (Gardner, 1990) is seen in a rare longitudinal study by Benbow and Arjmand (1990) involving 1,247 persons initially identified in the seventh or eighth grade as "mathematically precocious". These individuals were observed again after they completed college to identify factors that contribute to high achievement in mathematics and the sciences. In addition to finding that a high SAT score at age 12 was a good predictor of subsequent performance (however, a mediocre or low score did not yield much predictive value), the authors (Benbow and Arjmand) confirmed the observation of Walters and Gardner (1986a) that there was typically some "crystallizing experience" (event or persons) that contributed to the educational development of the high achievers (p. 437). Two observations are suggested here: first, that testers cannot rely on negative evidence as much as positive evidence of abilities, and second, that influence stemming from interpersonal relations (a mentor or encourager) may have a profound influence on mathematical or scientific achievement. Now this last outcome would seem to be excessively unlikely if the separate "intelligences" labelled "interpersonal" and "logical-mathematical" were truly quite independent. They have to be related via some form of intertranslatability.

The semiotic model under consideration (Figures 3-5) also enables us to make certain distinctions that are, it would seem, critical to any theory of intellect that aims for explanatory adequacy (cf. Chomsky, 1965). For instance, we may distinguish innate from acquired knowledge. Innate knowledge is that which is present before any experience occurs, or which is triggered by experience and matures more or less automatically and somewhat independently of experience. Even sensory-motor systems have their noteworthy conventional aspects. For instance, to take a trivial but suitable case for the sake of illustration, in one culture it is customary for automobiles to drive on the right hand side of a roadway while in another motorists stay to the left. If it is hypothesized that conventional aspects of the various semiotic systems in question must be acquired, this sort of acquired knowledge will be distinguished from innate knowledge to the extent that the former is a product of experience involving the senses. It is suggested that information from the sensory-motor system passes to consciousness where the sensory-motor texts (i.e., sequences of sensory-motor images) are interpreted. As they are understood, and just to that extent, they are passed through various stages of memory more or less distant from consciousness. The depth of the comprehension in question will determine the degree of impact on semiotic systems. It is hypothesized that the acquisition of grammar is a process of comprehending a particular kind of texts so as to develop the sort of intuitive feel which constitutes knowledge of a language. By this reckoning, the acquisition of a particular grammar is a process of comprehending texts in that language at a sufficient depth so as to acquire the conventional aspects of the grammatical system.

Contrary to a lot of recent speculation about non-primary language acquisition (e.g., Gregg, 1988), the theory under consideration hypothesizes that non-primary language acquisition will proceed in a manner much like primary language acquisition except for the fact that acquisition of a second language will benefit greatly (and suffer minor interferences from) the prior acquisition of the first language (Asher, 1969; Asher and Price, 1967; Asher and Garcia, 1969). Similarly, the acquisition of a third language will benefit (mainly, and suffer but little) from the first and second, and so on. The fact that non-primary language acquisition usually falls short of the mark achieved in primary language acquisition (Gregg, 1988), it is supposed, should be explained not by positing a radical difference in the physiology (Scovel, 1988) or even the internal strategies of the person involved in one or the other task (Selinker, 1972), but by noting the radical differences across the two cases in access to target language texts and the relative motivations to comprehend and produce them (Brown, 1973; Schumann, 1975; Vigil and Oller, 1977).

In the primary language situation, the person doing the acquisition is under incredible community pressure to conform to the norms of the primary-language. A child who persists in non-conformities will be ostracized or punished in ways that border on cruelty while the one who succeeds in overcoming them will be rewarded by all the privileges of membership in a community. For any one other than a child acquiring a non-primary language, no similar pressures or rewards are likely to be experienced (cf. Brown, 1973; Schumann, 1975; Vigil and Oller, 1976; etc.). Exceptional cases, where non-primary language acquisition succeeds in fairly dramatic ways are precisely those cases where access to target language texts and susceptibility to pressures and rewards are both provided for. For instance, the person who marries across language boundaries and then moves to the country where the non-primary language predominates is far more apt to achieve native-like ability in the non-primary language than someone who merely takes a college course in that language. In fact, we are inclined to suppose, along the lines of Vigil and Oller (1976) that continuing progress toward native competence in any language is much more a function of internally defined motives and sensitivities than it is a function of methods of teaching or modes of exposure. Clearly access to pragmatically rich and meaningful texts in the target language is requisite, but insufficient by itself. Motivation to conform to the communal conventions of the target language system is also required.

The hierarchical model under consideration not only supports the kinds of theoretical distinctions that are required in practice, e.g., the distinction between innate and acquired knowledge, consciousness and memory, memory and grammatical knowledge, grammar and text, text and comprehension, comprehension and production, primary and non-primary language acquisition, etc., but it also suggests some fairly explicit hypotheses about relationships within the proposed hierarchy that are immanently susceptible to empirical testing.

Since linguistic representations are the most abstract ones considered in the model, it follows that the primary language is the most likely basis for the development of general semiotic capacity. Here I differ some with Walters and Gardner (1985, 1986a, 1986b). They seem to view "logical-mathematical intelligence" as distinct from "linguistic intelligence." But, it has often been observed that logic and mathematics involve kinds of reasoning that are parasitic and derivative being entirely dependent upon language (Pierce, in Hartshorne and Weiss, 1931-1935; Lotz, 1951; Church, 1951; Russell, 1919). Einstein alluded to the closeness of the relationship between language development and cognitive growth in general in the remarks quoted above. It was a point developed further by Vygotsky (1934, 1978), Piaget (1947), Luria and Yudovich (1959) and Luria (1961).

Further evidence may be seen in the remarkable accomplishments of deaf children with hearing parents. In cases where the children, for whatever reasons, are deprived of access to visual sign language they face a language acquisition problem far more difficult than that of the hearing child. Such children, it seems, face special cognitive difficulties that only the acquisition of a fully developed language system will enable them to overcome. Typically this is accomplished through a natural visual-manual sign system such as American Sign Language (cf. Lane, 1984; Wilcox, 1988). (An interesting aside concerning such signed systems is that the primary role of language is assumed by gestures of the hands and body while the paralinguistic role of kinesics is taken over by speech and voice mechanisms.) Deaf children deprived of manual/visual sign system and forced to acquire speech directly are placed at a serious disadvantage (Lane, 1988). The difficulties they face in cognitive development across the board are predicted by the hierarchical model under consideration. It follows that if children are deprived of full and rich primary language system that is accessible to them in terms of their sensory-motor system, they will suffer consequences of this lack throughout the cognitive hierarchy and especially in areas that depend on communication, e.g., social development.

Moreover, children who acquire some ASL and are then taught Signed English (SE), an artificial system invented by hearing persons to correspond to English lexicon, syntax, and so forth, are apparently in the position of persons trying to acquire a second language system. In this instance, however, the system is artificial in a variety of ways. For instance, in theory SE gives equal emphasis to stressed and unstressed morphological and lexical elements. In this respect, and others, it is somewhat like Morse Code or even Pig-Latin. Unlike ASL, SE is a largely dependent system. Therefore, when deaf children de-emphasize or omit redundancies of English structure, e.g., the "-ing" of present progressives and the like, they are making natural modifications in surface forms of signed texts that would conform to more normal expectations about universal grammar.

Another hypothesis that is suggested by the theory under consideration is that neighboring elements of the hierarchy are more apt to influence each other than distant ones. For example, the primary language would have greater impact on second language acquisition than on third. The second similarly would be expected to influence the third, even more than the first language would, and so on. Again, experience of polyglots bears this out. Typically, "padding" (a term from Newmark, 1966, i.e., the use of known language forms in place of target language forms) is usually from the most recently acquired language rather than from any other.

Following out the same idea, transfer in general would be expected to occur from the more developed systems to less developed ones. For example, the primary language would be expected to influence a non-primary language rather than the reverse. The situation would be altered in favor of the non-primary language at just the point where the person in question achieved greater proficiency in the non-primary system. However, at just that point, the non-primary system would be promoted to the status of the primary system and the former primary system would presumably be demoted to a secondary status.

Another consequence of the postulated hierarchy is that distinct representational systems provide the means in some cases for comprehending what would otherwise be incomprehensible. For instance, a discourse in a target language that might be entirely incomprehensible if one had to rely on knowledge of that particular language alone can be made comprehensible if one has access to a translation provided in some other semiotic system. In normal language acquisition, e.g., primary language acquisition, as has often been pointed out (Macnamara, 1973, 1982) meanings of surface forms are often contextually obvious when those forms are being acquired (Krashen, 1985). The child first understands the context, e.g., by representing it in a comprehensible sensory-motor form, and subsequently becomes able to understand the utterances associated with the context. In non-primary language acquisition, wherever it succeeds, a similar scaffolding is often provided. It may be presented in some dramatization, in a film, or it may be presented through a translation, literally, into a language that the subject already knows.

By this line of reasoning, Krashen's input hypothesis (Krashen, 1985) is vindicated (Oller, 1988). The input hypothesis in its most basic form says simply that language acquisition progresses as the acquirer comprehends texts that are a little beyond his or her current level of development in the target language. Spolsky (1985) and Gregg (1988) have contended that the input hypothesis is either false or trivially true. If it means we must understand what is beyond our understanding, it is false. If it means merely that we must comprehend in order to learn, it is trivially true. However, the theory we are advocating here disposes of both of these interpretations. We do indeed understand representations (target language texts) beyond our reach in one system (namely the target language) by appealing to representations in another semiotic system. The one provides an interpretation of the other. Therefore, because of the intertranslatability of semiotic representations, the input hypothesis remains viable.

Cummins (1976) proposed the threshold hypothesis, an idea that relates to the impact of bilingualism, or more specifically adding a second language, on cognitive development. Subsequently (see Cummins, 1984, pp. 107-108) he modified his hypothesis and extended it. The threshold hypothesis suggests that the child's starting level of proficiency in one or both languages may be an important mediating variable in avoiding a burden in becoming bilingual or in benefitting from bilingualism once achieved. There are actually two thresholds being proposed.

On the low end, it is claimed that a child may have to achieve a certain minimal level of proficiency in one or both languages in order to avoid deficits. In other words, if the child falls below threshold in both languages, presumably it will be difficult or even impossible for that child to benefit from instruction in either language. Further, it follows that a child who has not acquired threshold level in the primary language will only receive an unnecessary additional burden by being instructed in a second language. Therefore, the lower threshold is presumably important in the determination of when instruction might be beneficially introduced in a non-primary language.

At the other end of the scale, a high threshold is also posited. In order for a bilingual child to experience the expected benefits of bilingualism, e.g., greater ability to appreciate and utilize symbols and greater "metalinguistic awareness," i.e., ability to appreciate the arbitrariness and conventionality of linguistic symbols, the child must have surpassed the high threshold presumably in one or both languages.

Admittedly, the idea of one or more thresholds is loosely stated, but the research seems to support it (Cummins and Mulcahy, 1978;

Duncan and DeAvila, 1979; Hakuta and Diaz, 1984; Kessler and Quinn, 1980). In fact, as Hakuta (1986; also see Lambert, 1975) has shown, there is a long history of debate concerning the deleterious versus beneficial effects of bilingualism. Formerly, especially in the U. S. there was a widespread prejudice against "bilingualism" based on research showing that minority language children got low scores on IQ tests. It scarcely occurred to the persons interpreting the research that the IQ tests were mainly measures of English language proficiency -- something that the minorities in question had not yet had the opportunity to acquire.

The main point here, however, is that the hierarchical model under consideration explains the available evidence concerning the threshold hypothesis and provides a convenient framework within which to understand the interrelationships of semiotic systems in general. Within a hierarchical model, the threshold hypothesis can be incorporated and elaborated in terms of transfer and interference and in terms of a more explicit theory of the role of language proficiency in relation to cognition in general. Bilingualism and indeed multilingualism deserve special consideration since they are bound to play a central role in the education of minorities. Moreover, the elaboration suggested by the theory under consideration is compatible, it seems, with the course that Cummins (1979, 1983a, 1983b) has begun to develop in terms of the CALP/BICS distinction.

In response to consideration of the possibility of a general language proficiency factor, Cummins (1979) hypothesized a distinction between what he called cognitive academic language proficiency (CALP) and basic interpersonal communicative skills (BICS). This idea was appealing inasmuch as most any educator who has dealt with bilingual or multilingual contexts has observed ample evidence in its favor. A child that gets along satisfactorily on the playground, where cognitive demands are presumably lessened by the immediacy of physical and social context, may encounter difficulty in the classroom when it comes to reading, writing, solving word and math problems, and in general interacting on a more abstract level. The child may have adequate BICS without sufficient CALP. This distinction is reminiscent of the sort of thing Gardner (1990) says in reference to representational systems that seem to be naturally acquired versus ones that need special "tutelage" -- especially, "literacy, numeracy, and critical thinking" -- the sorts of things that Cummins would group under CALP. Cummins (1983c), however, unlike Gardner and colleagues, clarified that he did not intend to argue that the two kinds of ability were unrelated, but rather that they were apt to appear as such at the surface. To illustrate he adapted an "iceberg" model (from Shuy, 1978, 1981) where the two visible points, CALP and BICS, were clearly distinct, but were joined below the surface in what he called "common underlying proficiency" (cf. Cummins, 1984, p. 143).

There was a further implication that the two kinds of ability might be developed in somewhat different contexts and perhaps using distinct strategies. Cummins (1983c) quoted David Olson (1977) who said:

> ...language development is not simply a matter of progressively elaborating the oral mother tongue as a means of sharing intentions. The developmental hypothesis offered here is that the ability to assign meaning to the sentence per se [as in a written text], independent of its non-linguistic context, is achieved only well into the school years (p. 275, cited by Cummins 1983c, p. 116, our interpolation).

What Cummins and Olson apparently intend to emphasize is the greater degree of inference required to link up a written text with its author's intended meanings than is required in the case of an interactive discourse in the here and now. The latter, presumably the typical context of the exercise of BICS, is less cognitively demanding, *ceteris paribus*, than the former, a typical context for the use of CALP.

Within the more elaborate Peircean perspective proposed here, Olson's phrase "independent of its nonlinguistic context" might be reformulated as "without firsthand access to its nonlinguistic context." This seems to do no violence to Olson's intention, nor Cummins application of the idea in reference to CALP. However, it is a necessary modification if Pierce's foundational claim that all interpretation is translation from one form of semiotic representation to another. This sort of translation is not viciously circular only because sensory-motor representations enable the investment of all other sorts of representation with material (non-empty) content.

However, strictly speaking, there is no such thing as a meaningful "sentence" without a "nonlinguistic" context. With that in mind, we assume that Olson and Cummins might accept as a friendly amendment to their ideas the interpretation that CALP (or in Olson's case, literacy) requires a larger inferential leap from the perceptible form of a representation (a written text in the case under consideration) and an appropriate interpretation that associates it with experiential context. Failing this, it would have to be argued that a representation which has no inferential relation to any experiential context whatever is necessarily meaningless. It is entirely uninterpretable (cf. Einstein, 1944, in Oller, 1989, p. 25, paragraph 3.13; and Pierce, pp. 99-105 in Oller, 1989).

How then can the CALP/BICS dichotomy be understood within the proposed hierarchical model? The overlapping part of the iceberg beneath the surface would be explained in part as the general factor of language proficiency which incorporates whatever aspects of general intelligence are necessary to that proficiency. For BICS, also, it is clear that the utilization of both sensory-motor information and linguistically coded representations simultaneously would require a pragmatic linking that could only be accomplished by access to general semiotic ability. However, with BICS, sensory-motor information is immediately accessible to aid the pragmatic linkage.

In the exercise of CALP, on the other hand, say in reading an unillustrated text, e.g., that which appears on this page, any necessary supplementary sensory-motor representations would have to be supplied by the reader. This is a more difficult semiotic task. It requires a higher degree of inference based on a more abstract semiotic system, namely a linguistic one, from which the sensory-motor type images must be inferred where they are needed. The move from graphological representations to a more abstract linguistic form is already a difficult inferential process (reading), and the absence of sensory-motor images that might give some clue concerning reference, deixis, and the whole pragmatic mapping process involves another complex of inferences.

Thus, CALP, with its special emphasis on literacy and abstract reasoning would presumably require the development of reading and writing skills in the primary or some non-primary language. Whereas BICS might benefit indirectly from such a development, literacy and specialized abstract reasoning skills, e.g., ability to do arithmetic leading on to higher mathematical skills, would not be necessary to BICS. To this extent, BICS and CALP are usefully distinguishable which suggests an important amplification of Cummins's threshold hypothesis -- one that he has commented on (Cummins, 1984, p. 117).

The initial distinction between "surface fluency" and "conceptual-linguistic knowledge" Cummins attributes to Skutnabb-Kangas and Toukomaa (1976). They, no doubt, were influenced by the distinction between "surface" structure and "deep" structure from Chomskyan linguistics. The idea was that a child might develop quite a lot of routine facility with greetings, leave-takings, playground games, and the like, and still fall short of the level of language proficiency and concept development necessary to reading, writing, and doing arithmetic (or as Gardner, 1990, terms them "literacy", "critical thinking", and "numeracy"). Therefore, a child might appear to do well at conversation but fail at school (Olson, 1977).

The low threshold for language skill, then, might be construed as a completely general requirement applying as much to monolinguals as to multilinguals. Presumably this same notion was what another generation of specialists in another paradigm meant by "readiness". The higher threshold too would have a more general interpretation in this context. Presumably "metalinguistic awareness" is merely another way of referring to what another generation of psychologists and educators called "learning to learn" or "talking about talk," etc.

Finally, there is also a parallel with the traditional distinction between "language disorders" and "learning disabilities" where the former have been defined more in terms of surface language problems (sometimes even speech difficulties per se) and the latter in terms of deeper conceptual difficulties -- "neurological" deficits (see Coles, 1978; Cummins, 1986) or, more recently, "inefficiencies" (Swanson, 1988). Damico (1985b) has argued that traditional tests of language disorders have tended to focus on surface forms of language while definitions of learning disabilities have been defined, to the extent they have been defined at all, in terms of deeper conceptual problems. Again, something like the BICS/CALP distinction appears. It is a virtue of the proposed model under consideration to be able to incorporate such distinctions and to elaborate upon them in intuitively appealing ways.

Table 1
The Seven Intelligences

Intelligence	End-States	Core Components
Logical-mathematical	Scientist Mathematician	Sensitivity to, and capacity to discern, logical or numerical patterns; ability to handle long chains of reasoning.
Linguistic	Poet Journalist	Sensitivity to the sounds, rhythms, and meanings of words; sensitivity to the different functions of language.
Musical	Composer Violinist	Abilities to produce and appreciate rhythm, pitch, and timbre; appreciation of the forms of musical expressiveness.
Spatial	Navigator Sculptor	Capacities to perceive the visual-spatial world accurately and to perform transformations on one's initial perceptions.
Bodily-kinesthetic	Dancer Athlete	Abilities to control one's body movements and to handle objects skillfully.
Interpersonal	Therapist Salesman	Capacities to discern and respond appropriately to the moods, temperaments, motivations, and desires of other people.
Intrapersonal	Person with detailed, accurate self-knowledge	Access to one's own feelings and the ability to discriminate among them and draw upon them to guide behavior; knowledge of one's own strengths, weaknesses, desires, and intelligences.

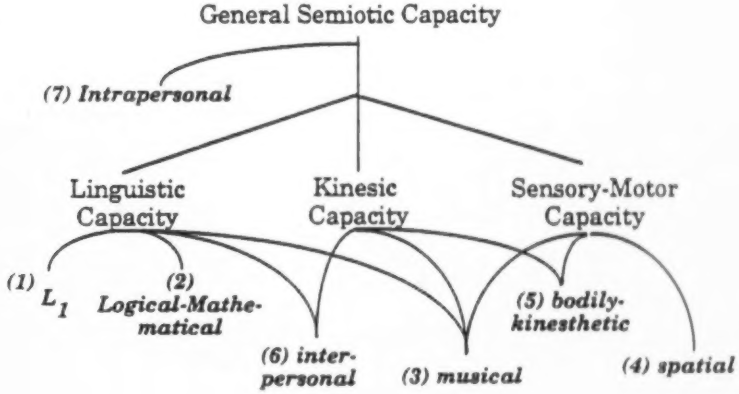
To see better how the proposed hierarchy works in practice, and also to show how it can be used in the evaluation of other theories of intelligence, it may be useful to pause to examine more closely the model proposed by Gardner (1983, 1989, 1990) and colleagues (especially, Gardner and Hatch, 1989; Walters and Gardner, 1985, 1986a, 1986b). Table 1 gives a list of the seven "intelligences" that Gardner sees as somewhat independent of each other and yet as capable of characterizing of the sorts of individual configurations of abilities that he believes necessary to a more adequate conception of intelligence. While Gardner and colleagues speak as if their categories of "multiple intelligences" were thoroughly independent, they are upon examination hardly self-contained, independent modules, but rather complex composites of semiotic capacities in each case. Perhaps they are quasi-modular in character, but it is difficult to see them even in that way. Nevertheless, for the sake of demonstrating the intrinsic compatibility of the quasi-modular semiotic hierarchy I have been discussing here (Figure 4 above especially), I will fit Gardner's categories in as shown in Figure 5 and will discuss them one-by-one in terms of the analysis given by Gardner and Hatch (1989) as well as my own semiotic characterization of their categories.

The first category is what they call "logical-mathematical intelligence" which they describe (see Table 1 above) as pertaining to a "scientist" or "mathematician". It is generally agreed by professional logicians and mathematicians (who have gained some awareness of linguistics) that logic and mathematics are both parasitic and derivative fields of study entirely dependent on human language abilities at a deep level. Therefore, I have placed Gardner's first "intelligence" as a node subordinate to the universal deep language system that is postulated to underlie all abstract symbolic systems as well as natural languages.

Gardner's second category, "linguistic intelligence" characterized in the special proclivities of a "poet" or "journalist" I have associated with primary language ability in the semiotic hierarchy. Gardner and Hatch (1989) give no indication that they have in mind any sort of polyglot, so I do not relate their category directly to the deeper level of universal language ability. That deeper level, I suppose, must undergird all abstract symbol systems such as mathematics, logic, and musical notation, as well as the abstract symbolic aspects of map making, diagramming, illustrating, and in general all forms of what Pierce called "abductive reasoning" (or what I term "pragmatic mapping"; as diagrammed in Figure 3 above).

Gardner's third category, "musical intelligence," as shown in the special abilities of a "violinist" or a "composer," I would place under the sensory-motor class of representations but with special connections to deep language abilities and to kinesic abilities. While a violinist might not be a reader of musical notation, this is unlikely, and a composer certainly would be a reader of music -- hence the connection with the abstract deep language node. In addition, a composer or a violinist would also be apt to understand the sorts of special gestural systems used by conductors (though neither of them might be conductors, a composer would be likely to have the capacity to conduct one or more musicians in performing his or her music) -- hence, the connection with the kinesic (significant gestural) node.

Figure 5
The Semiotic Hierarchy with
Gardner's Seven Categories ("Multiple Intelligences")
Added to the Picture


General Semiotic Capacity
(7) Intrapersonal
Linguistic Capacity
Kinesic Capacity
Sensory-Motor Capacity
(1) L_1
(2) Logical-Mathematical
(5) bodily-kinesthetic
(6) inter-personal
(3) musical
(4) spatial

The fourth kind of intelligence, "spatial," as represented in the special skills of a "navigator" or "sculptor" seems remarkably broad. Surely it covers a multitude of abilities. Among them would have to be found the sensory-motor elements pertaining to perspective and movement in time and space as well as a keen sense of proportion bordering on the mathematical. For the navigator, mathematical skills would surely come into play. For this reason, the "spatial intelligence" is connected both to the sensory-motor node and to the deep language node.

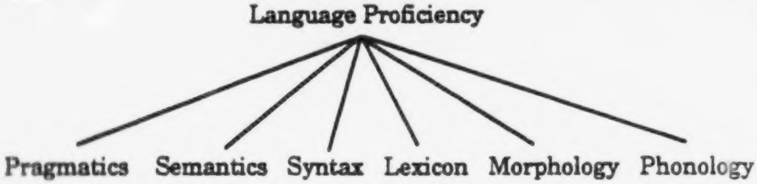
"Bodily-kinesthetic intelligence," Gardner's fifth kind of intelligence, as seen in a "dancer" or "athlete" suggests a multitude of connections as well. If the dancer is a person who understands choreography or if the athlete understands demonstrations of various performances (e.g., how to serve a ball in tennis or how to do a single-leg sweep in wrestling), an implicit comprehension of diagrammatic illustrations would probably come into play. Therefore, I have shown connections to the kinesic node as well as the sensory-motor, but no doubt if coaching comes into the picture, the language node should be connected as well.

The sixth category, "interpersonal intelligence" as seen in a "therapist" or "salesman" suggests again an interesting composite of abilities. Since "moods, temperaments," etc. (as suggested in the descriptor of the category) are discerned largely through kinesic and paralinguistic systems such as gesture, tone of voice, facial expression, and the like, the primary connection would be with the kinesic node. However, to the extent that all sales' pitches tend to rely on linguistic as well as other representations, at least the primary language system would come into play. Since Gardner and Hatch give no indication that the salesperson or therapist they have in mind is a multilingual, connections to languages other than the primary one are not shown, but a polyglot would no doubt have them. Therefore, it is clear that this module of "intelligence" would probably be heavily contaminated by one or more verbal components.

The seventh category is the most problematic of all. Gardner calls it "intrapersonal intelligence" and suggests that it is the ability to understand one's own abilities. The sort of person having this particular constellation of gifts is not only, we may suppose, a rare bird, but one who knows even more about him or herself than the people who are looking for him or her. That is to say, a person who understands his or her own abilities in the way described knows a good deal more than the measurement specialists do. This category, however, I suppose would have to be linked directly to the deepest level of the semiotic hierarchy since it implies knowledge of all the nodes beneath it and of their interconnections. This final observation concerning Gardner's system also sums up my basic objection to it: the interconnections that must be posited if we are to understand how the various modules relate are missing. The sort of semiotic hierarchy that I am proposing here, however, would supply at least some plausible alternatives for such connections.

One of the most difficult things to see about language proficiency is that it may (perhaps must or at least ought to) be conceptualized in a considerable variety of different but mutually compatible ways. Walters and Gardner (1985) assert that "a particularly high level of ability in one Intelligence, say mathematics, does not require a particularly high level of ability in another Intelligence, like language or music. This independence of Intelligences contrasts sharply with traditional measures of IQ that find high correlations among test scores" (p. 13). I agree in large measure with what they are saying provided we modify the word "independence" to "quasi-independence" or something of the sort.

Figure 6
Language Proficiency Viewed as a Composite of Domains of Grammar



With respect to language proficiency per se, it is possible to think in terms of the various components of grammar (Figure 6) that constitute it in theory, or we may think of language proficiency in terms of the traditional skills (Figure 7). Or, we may choose any number of other angles or combinations of them. What is difficult to see is that these are not incompatible ways of viewing the phenomena of interest -- merely different ways. If we focus on primary language ability as represented in Figure 4 above, that portion of the diagram might be amplified as shown in Figures 6 or 7. In Figure 6, language proficiency is seen as divisible, more or less, into domains of grammar. Pragmatics may be defined as pertaining to those aspects of meaning that have to do with actual, particular, concrete contexts of experience. Semantics embraces those aspects of meaning that are virtual, universal, or abstract. Syntax is concerned with the sequential or simultaneous arrangement of categories of grammar into texts. Lexicon comprises those inventories of elements that are acquired as whole units, e.g., words, idioms, pat phrases, verbal routines, and the like. Morphology in English is a question of inflections, e.g., pluralization, tense and number marking on verbs, etc., and derivations, e.g., adding a morpheme to make a verb of an adjective, e.g., "real" plus "-ize" to get "realize," and so forth. Phonology is a matter of determining the surface forms of phonemes, syllables, lexical items, and larger units of structure.

Figure 7
Language Proficiency Viewed as a Composite of Quasi-Independent Skills

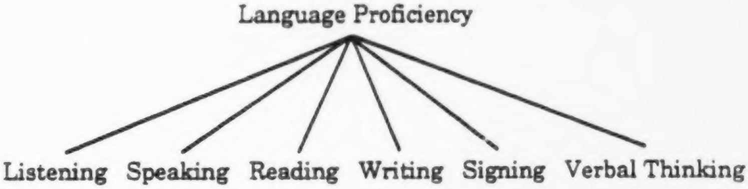


Figure 7 shows a similar breakdown with reference to skills such as listening, speaking, reading, writing, and verbal thinking. It may be argued without risk of contradiction that such hypothetical domains of structure, or distinct skills, are as valid as the theories upon which they are based. However, such divisions can never be finally determined anymore than Immanuel Kant could determine once for all the ultimate categories of reason. As Pierce, Einstein, and others have shown, such categories are intrinsically arbitrary and cannot be finally fixed or completely determined by any amount of empirical research (see especially Einstein, 1941, 1944, and Pierce, 1878, 1906). While it may be possible to fix upper and lower limits within which the simplicity/complexity of the model must fall, its specifics will apparently always retain a substantial arbitrariness nonetheless.

For instance, there is no conceivable argument that would prove either of the componential breakdowns of Figure 6 or 7 to be intrinsically superior to the other. For one purpose one model might be preferred, for some other purpose, another. What is more, many other componential models may be conceived. For example, modes of processing (productive versus receptive) may be distinguished, modalities of processing (articulatory/auditory versus visual/manual), stages of processing (consciousness, short-term, long-term memory), etc. In principle, there are an infinite variety of possible componential models. The answer, therefore, to the advocates of multiple intelligences (e.g., Gardner, Walters, and other collaborators) is that there is no single arrangement that will be completely satisfactory. Within the proposed hierarchy, this fact can be construed as a natural outcome of different ways of combining and/or parsing up various of the proposed elements.

While it was long maintained that cognitive development may be hindered by becoming bilingual, the evidence clearly points in the other direction (cf. Hakuta and Diaz, 1984; Cummins, 1984, 1986; Hakuta, 1986). Dabbling in non-primary language acquisition may have little or no impact on intellect, but the acquisition of a second or third or fourth language to a substantial degree of proficiency is apt to result in significant, though modest, cognitive gains. In particular, the evidence seems to suggest that bilinguals achieve some kinds of flexibility in reasoning and a capacity to appreciate certain kinds of abstract relations that might remain outside the reach of some monolinguals. This result (see the research cited above with reference to the "threshold" hypothesis), is predicted on the basis of the hierarchy under consideration.

Moreover, as in the case of the threshold hypothesis, a more general hypothesis is suggested. If bilingualism contributes to mental growth only after some threshold is passed, it follows that simply attaining proficiency in one's primary or native language must be important to normal mental maturation. Further, if language is a window through which researchers may get a fairly clear look at the mind, a thesis Chomsky has been pushing lately, it follows that the development of language proficiency must be linked to normal cognitive development. Putting this hypothesis in its most general form (Oller, 1991) following Pierce, Einstein, and others, it is possible to predict that the normal development of deep semiotic abilities must depend in subtle ways on the development of the primary language. This has been demonstrated above in part by the differentiation of iconic, indexical, and symbolic representations. Because of its greater abstractness (i.e., symbolic character), language has certain capabilities that the other representational systems lack. Among them is the potential for deep level semantic representations that are quite abstract (i.e., relatively uncontaminated by the two kinds of degeneracy associated with icons and indexes). As a result, only deep language ability is logically a medium that might serve for the development of the most general sort of intelligence. For an elaboration of this idea and a content analysis of so-called "non-verbal" IQ tests showing that they require such deep propositional or semantic reasoning, see Oller (1991).

While it may be possible for deep semiotic abilities to be developed to a high degree with reference to some other manifest form, say, sensory-motor representations, since linguistic representations achieve a more complete level of logical abstractness and conventional arbitrariness, it seems likely that in normal human beings language development in all of its diversity is the fulcrum on which intellect attains its greatest leverage. It also follows that language abilities will tend toward the center of any definition of human exceptionalities ranging from giftedness in all its varieties to disabilities of all types.

(4) Recommendations for Testing (and Teaching) LEP Students

Cummins (1986) writes, "Historically, assessment has played the role of legitimizing the disabling of minority students. In some cases assessment itself may play the primary role, but more often it has been used to locate the 'problem' within the minority student..." (p. 29). This process may not have been intentional, but the effect has been summed up by Chase (1977) in a single phrase. He called it "the biologizing of social problems" (cf. Coles, 1978, for concurrence).

Not to deny the fact that some children may indeed have genuine "neurological" or other "deficits" or even "abnormalities," Cummins still contends that the medical "diagnosis/prescription" paradigm has seduced a whole generation of educators and clinicians, and that in many cases children from minority language backgrounds have been ludicrously over-represented in deficit categories (e.g., see Ortiz and Yates, 1983). It is the purpose of this section to discuss these facts in light of the proposed model of semiotic abilities and to show some of the ways that the whole process of assessment might be upgraded and set on a path of self-correcting research and progressively greater adequacy.

It is difficult to over-estimate the pervasive influence of analytic, discrete-point thinking in the study of exceptionalities. Its main manifestation is the search for specific, particular, unique sources of difficulty in individual cases. Swanson (1988), for instance, stresses the aim of the learning disabilities paradigm to achieve "specificity" (p. 197) -- a concept that is elaborated throughout his informative article. This means focussing on "specific mental processes" in instructional remediation and determining unambiguously that "the process under investigation is responsible for performance" (p. 200). The idea of a "generalized deficit," he says, "undermines an important tenet of the field" (p. 197). He complains that "there is a lack of theoretical integration in the choice of measures in subtyping studies, and non-operational definitions of LD exist (Shepard and Smith, 1983). Further," he complains, "there is no agreed upon or satisfactory method for determining subtypes (McKinney, 1984)" (p. 197).

The demand, therefore, appears to be for more specific diagnosis and more specific remediation. These goals were characteristic of the discrete-point language theory of the 1960s in second and foreign language testing. Swanson (1988) shows that this same sort of thinking is current in the study of learning disabilities when he says, "Simply stated, a learning disability reflects a cognitive deficit...that is *reasonably* specific to a particular domain (e.g., reading).[3] The specific deficits displayed by such children must not extend too far into other domains of cognitive functioning. If they did, the concept of a learning disability would be meaningless..." (p. 196; his italics). However, Swanson goes on to observe that in fact "the literature has undermined the concept of specificity" (p. 197).

If we accept the major premise of Swanson that "the LD field is directed by social consensus" (p. 196), then it would follow that "the literature" which both establishes and defines the "consensus" could perhaps happily be redirected. However, I believe that it is not the "literature" per se that has "undermined the concept of specificity" as if there had been an active conspiracy against the "social consensus" that defines "the field of learning disabilities" (all the quoted terms being from Swanson, 1988). The evidence is simply against the idea of specificity in the way that it has been put forward. As argued extensively above, a more comprehensive and integrated view of semiotic capacities is needed to incorporate and explain rather than deny or purge the data of existing research.

A pragmatic approach, along the lines described above will be required, and the goal of isolating highly specific elements of cognition will generally have to be abandoned as a logical mistake. Cognition by its very nature involves the differentiation of specific elements only in rich and dynamic tensional contexts in which those elements find their distinctive identities. Apart from such contexts, those specific elements do not exist. This has been the primary motivation for clinical discourse analysis (Damico, 1985a, 1985b), an approach which seeks to understand the actual dynamics of the communicative performances of children rather than to pigeon-hole them into ready-made categories that may turn out to be altogether inappropriate in many cases. Discrete elements of cognitive processing only attain the character that really defines them in the contexts of their dynamic tensional oppositions in relation to each other and the whole continuum of experience (see the voluminous writings of Pierce on this matter as represented in collections by Burks, 1958; Hartshorne and Weiss, 1931-1935; Fisch, et al., 1982; Moore, et al., 1984; and Oller, 1989).

What about the current consensus that defines and purports to identify children with language disorders and/or learning disabilities? While the latter category has come more by tradition than by evidence to be associated with "neurological impairment", the idea that the former category is a subset of the latter is merely a matter of definition. The distinction between the larger category, learning disabilities, and the subcategory, language disorders (cf. Rueda and Mercer, 1985; also Cummins, 1986, p. 29), is merely assumed to be generally valid.[4] The distinction is never demonstrated by factual evidence anywhere in the vast literature on learning disabilities. In addition to a critical examination of this distinction, therefore, I wonder about the social consensus that sustains (Swanson, 1988) the whole field of special education and the study of exceptionalities in general.

As soon as the National Advisory Committee on Handicapped Children (1968) launched the first sentence of its long-standing definition of "learning disabilities" the confounding of that term with "language proficiency" and therefore with "language disorders" should have been abundantly apparent. From there forward, the problem of providing a theoretically adequate basis for the sought after distinctions only becomes more confused. They wrote:

> Children with learning disabilities exhibit a disorder in one or more of the basic psychological processes involved in understanding or using spoken or written languages. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, or arithmetic. They include conditions which have been referred to as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, developmental aphasia, etc. They do not include learning problems which are primarily due to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to environmental disadvantage (p. 4).

What is remarkable is that a vast number of workers could be encouraged to entertain the illusion that the kind of thinking expressed by the NACHC (and similar bodies) was a sufficient foundation on which to erect the present superstructure of the vast and growing edifice of special education.

Coles (1978) reviewed ten of the most widely used procedures for identifying children with the sorts of "disabilities/disorders" supposedly defined in the previous paragraph. He examined the *Illinois Test of Psycholinguistic Abilities*, *Bender Visual-Motor Gestalt Test*, *Frostig Developmental Test of Visual Perception*, *Wepman Auditory Discrimination Test*, *Lincoln-Oseretsky Motor Development Scale*, *Graham-Kendall Memory for Designs Test*, *Purdue Perceptual Motor Survey*, *Wechsler Intelligence Scale for Children—Revised*, neurological evaluations, and electro-encephalograms. These were found to be the most common procedures in use for the identification and diagnosis of learning disabilities in most states.

The sad conclusion was that "the predominant finding in the literature suggests that each test fails to correlate with a diagnosis of learning disabilities" (p. 326). Neither was there evidence of correct diagnosis in the results of therapeutic interventions: "In experiments where the dysfunction itself was treated, there was little success" (p. 326). While correlation alone is never proof of a causal relation, the absence of correlation is fatal to theories about specific causal connections. At the end of his article, Coles asserted, somewhat optimistically it would seem in retrospect, that "there is little question that eventually the tests reviewed here will be discarded; the evidence against them is mounting" (p. 335). If we think in terms of centuries rather than decades, this statement may yet turn out to be correct. At the moment, the tests in question are probably being used in about as many states and in far more cases in 1991 than they were in 1978.

When it comes to the subset of learning disabilities known as language disorders, there is even more confusion, if that is possible. The deep underlying question is what do tests used to define language disorders (and learning disabilities) really measure? The theory is that they should measure something over and above whatever intelligence tests measure. According to most researchers they are supposed to identify actual "neurological impairments" or at least "neurological inefficiencies" (Swanson, 1988).

However, if we take a paradigm exemplary test such as the *Illinois Test of Psycholinguistic Abilities* (Kirk, McCarthy, and Kirk, 1968), it turns out to be notably ineffective in predicting even reading scores if we control for IQ. Newcomer and Hammill (1975) reported that the correlation between ITPA scores and reading scores evaporated when intelligence was used as a covariate. Our point here is not to defend IQ tests as such (on the contrary, see part B below), but to show how confounded the constructs of language disorders, learning disabilities, and IQ are with each other. Moreover, we are arguing that all of these constructs have tended to overlook what is probably the single most important mediating variable, namely, primary language proficiency.

In general there has been a consensual distinction between "mental retardation" and "minimal brain damage" or "neurological impairment." Mental retardation is supposed to be related to, among other things, scores below some arbitrarily established level on standardized IQ scales. We, like Cummins (1984, see note 9, p. 30), do not deny that brain damage occurs in some cases or that mental retardation is in some instances a useful designation. What we do question, on the other hand, is whether these categories can be and are adequately distinguished on the basis of the present approach to IQ measurement and learning disabilities diagnosis (also see Mercer, 1973; Briere, 1973). There is substantial evidence that the distinction is thoroughly confounded in large numbers of cases. For instance, children identified as having "learning disabilities" in many cases are well below average in IQ scores. Out of 3,000 "learning disabled" children (identified as such in twenty-one states), more than a third fell below 90 on the standard IQ scale (Kirk and Elkins, 1975).

Why would educators tend to place at least some "mentally retarded" cases in the "learning disabled" category? It is clear that the former category is more stigmatized than the latter, and that the compassionate diagnostician, psychologist, or whatever, will prefer the less damaging label. But the problem surely runs much deeper than this. Beers and Beers (1980) point out that in some school systems a fourth to a third of the total school kindergarten population is being flagged as "potentially" learning disabled. This seems odd when a dramatically smaller percentage of the population is apt to have either genetic or acquired physical disabilities. Cummins (1984) aptly describes the category of "learning disabled," therefore as "a dumping ground for a wide variety of learning and behavioral difficulties" (see also Hallahan and Cruickshank, 1973). Swanson (1988) confesses that there is not a single trait, nor even a cluster of them, that can be identified as common to the category.

Undoubtedly it was because of the profound degree of confusion about the relation between mental retardation and learning disabilities that the American Association of Mental Deficiency arbitrarily changed the definition of "mentally retarded" from one to two standard deviations below the mean on a standardized IQ scale (McKnight, 1982). Cummins (1984, p. 83) sees this change as motivated by the desire to reclassify large numbers of formerly "mentally retarded" children as "learning disabled." A question that immediately arises is what such a change means in reference to the underlying constructs of intelligence versus neurological impairments. Beyond this, there is the lingering question of how language proficiency may be construed as relating to either of these constructs. What is disturbing is that in their educational applications both constructs are becoming, it would seem, increasingly folkloric and arbitrary.

Traditionally the identification of children with "language disorders" or "communicative disorders" or the general run-of-the-mill class of "learning disabilities" has been based on fairly superficial, surface-oriented criteria. For example, traditional diagnosticians have asked whether or not a child appropriately uses plural nouns (e.g., "dogs" versus "dog"), possessives (e.g., "Jim's hat" versus "Jim hat"), third person singular non-past verbs (e.g., "he walks" versus "he walk"), past tense verbs (e.g., "wanted" versus "want"), noun-verb agreement (e.g., "I am" versus "I be" or "I is"), irregular verbs (e.g., "fell" versus "falled"), number concord (e.g., "these cats" versus "this cats" or "these cat"), auxiliaries (e.g., "they have gone" versus "they gone," "they be gone," or "they done gone"). With respect to phonology, clinicians have tended to emphasize such things as the various forms of the regular plural morpheme in English (viz., /-z/, /-s/, or /-^z/) and the similar variations that occur in possessive marking of nouns, the third person singular non-past marking of verbs, the contractions of "is" and "has," and the similar variations that occur in marking of regular past-tense verbs (viz., /-d/, /-t/, or /-^d/).

Of course, surface form has some significance in its own right, but it has been elevated in the traditional tests, measurements, and diagnostic procedures of speech-language pathologists to such a position of prominence that the deeper purposes, the pragmatic aims of communication have been overlooked. As a result, "language disorders" have typically been defined in terms of superficial elements of syntax, morphology, and phonology, and more often than not have been strictly limited to problems of speech and writing rather than deeper aspects of the production and comprehension of meaningful discourse. Not only has the diagnostic definition of "language disorders" qua "learning disabilities" been based on surface-oriented criteria traditionally, but the treatment of them has likewise focussed on "intensive instruction in phonics" and "perceptual training" (cf. Beers and Beers, 1980, p. 73). The remedies, like the diagnoses, have been largely ineffective (Coles, 1978).

When attention is turned to discourse processing and to pragmatic criteria that have the potential at least of tapping into the deeper conceptual processes that underlie it, it is expected that the identification of genuine communicative difficulties, the kind that are apt to influence academic achievement in dramatic ways are more apt to be turned up (Damico and Oller, 1980; Damico, Oller, and Storey, 1983; Damico, 1985a, 1985b; Damico and Oller, 1986; McCord and Haynes, 1988[5]). This is not to say that researchers are presently in a position to determine on the basis of any existing testing program the specific neurological correlates of a given performance. This may be possible in rare cases but is certainly not the norm. Rather, as Coles (1978) intimated, there are no fully developed "less well-known instruments standing in the wings" (p. 335) and ready to fill the present void of thoroughly validated diagnostic procedures. As Coles said, "These tests, in any case, do not yet exist" (p. 335), and even the theory for their development is largely lacking.

What chiefly stands in the way of the needed theoretical and practical development is the uncritical acceptance of the present "social consensus." If researchers and practitioners alike are willing to acquiesce to the status quo of existing categories such as "language disorders," "learning disabilities," "mental retardation," and in general to the whole "diagnosis/remediation" paradigm, the needed reform of theory and practice is bound to be delayed if it ever comes at all. As Cazden (1985) has argued, the labeling of minority children especially as "disabled" or "disordered" must be, in her words, "delegitimized" and this can only be accomplished by looking to the broader context of socialization and education as has been argued by Coles (1978), Cummins (1984, 1986), and by Oller and Perkins (1978).

Based on all of the foregoing a few heuristic guidelines may be offered. Since the damage is likely only in cases of disabilities rather than giftedness, we concentrate on the former. To begin with there are logically just four types of errors to be avoided: (1) a LEP may be wrongly identified as disabled; (2) a truly disabled LEP child may be left out of the disabled category; (3) a LEP child may be incorrectly classed as a non-LEP; or (4) a non-LEP may be classed as a LEP.

It is known that large numbers of errors of type (1) are occurring. Many LEPs are incorrectly being diagnosed as disabled, or otherwise retarded. It follows from the same studies documenting type (1) errors that type (2), disabled LEPs not being identified as such, must also be common. Error type (3), LEPs incorrectly classed as non-LEPs, seems most likely when in Cummins' terms a child has developed substantial BICS in English but not much CALP. In these cases educators are apt to be fooled into thinking the child is ready for literacy in English when the child is still below threshold even in his or her primary language. Error type (4), non-LEPs classed as LEPs, can also occur if the child is evaluated on the basis of limited BICS while well-developed CALP in the child's primary language may be overlooked. The likelihood of a growing number of misclassifications of all four types is on the upswing due to the increasing number of non-English speaking minorities in our schools.[6]

To minimize errors of all four types a series of assessment phases is recommended. In all phases, the pursuit of evidence concerning the child should be treated in a matter-of-fact manner and with a view to the advocacy of the interests, needs, and feelings of the child above those of the school or the diagnostician. Our purpose as educators should be to promote and guard the interests of the child, not those of some abstract political or educational entity such as a state, institution, profession, or psychological yardstick (Cazden, 1975; Coles, 1978; Cummins, 1986).

First, to distinguish LEPs from non-LEPs, a variety of sources of evidence should be considered, e.g., talk with the child, observe the child's behavior in casual contexts, talk with siblings, parents, friends, etc., where appropriate. Ask about literacy and previous educational experience. Keep in mind that superficial, routine verbal skills may be deceptive in two ways: (a) they may lead us to attribute more language ability than is really present, or they may seem to indicate a low level of academic readiness when in fact the child is already literate in one or more other languages. Clear-cut cases may be decided on the basis of this preliminary phase to be either LEP or non-LEP. Doubtful cases should be referred to the second phase of assessment.

Two kinds of doubtful cases may be distinguished. Children with substantial educational background, e.g., those who have attained literacy in one or more other languages, but who lack basic routine skills (BICS) in English constitute the first case. These children should be evaluated with reference to their attainment in their most developed or primary language(s). For instance, some Asians will prove to be weak in English but literate in French and possibly some other language. To determine this fact may require additional interviews and possibly testing in the primary language. The question to be addressed in these cases is presumably, would it best serve the interests of this child if he or she were mainstreamed? If Cummins (1984 and elsewhere) is correct in the threshold hypothesis, only children who have demonstrated fairly advanced literacy skills or other abstract linguistic capabilities should be mainstreamed.

The other kind of doubtful cases referred from phase one would include the children who appear to have substantial ability to perform routine tasks in English (BICS) but who may or may not be ready for academic mainstreaming. The determination here, as in all cases, should be based on the solution that is believed most likely to benefit the child optimally. Preferences on the part of the child, and or the child's parents, should be weighed together with further evidence concerning academic readiness. The latter should be evaluated mainly in terms of the child's ability to perform abstract reasoning in the primary language and/or in English. Again, if Cummins (1984) is on the right track and if the theory as discussed above is followed in a general way, well-developed abstract reasoning capacities in one language will easily transfer to another assuming that there are no affective or social barriers[7] actively interfering with the process. In short, presumably some of these children should be mainstreamed, and some should not.

Phase three concerns children who have been identified as LEPs needing some kind of special program to enable them to profit optimally from their on-going educational experience. The objective during this phase is to differentiate children who are ready for a normal course of instruction in their primary language and those who may need some extra help beyond this. The latter are those traditionally labeled "learning disabled."

At this point, teachers or competent para-professionals who know the primary language(s) of the children should have already been involved and now become the main assessors. They should be trained in the deeper kinds of language assessment procedures that look to discourse/text-based tasks that include the broad range of communicative activities that school children are becoming able to engage in: e.g., relating an experience, singing a song, reading and reacting to a story, drawing a picture to illustrate some idea, explaining an illustration, evaluating a facial expression or gesture in a filmed narrative, play or drama, writing a letter, answering an advertisement, etc. The list of tested activities should be as broad as the curriculum children are expected to cope with. As suggested by Damico, Oller, and Storey (1983) and elaborated by Damico and Oller (1985) as well as Damico (1985a, 1985b, 1991) LEP students should be assessed in all of their languages and in each case across the broad spectrum of abilities so as to identify strengths. The objective at all points along the way should be not to look merely at surface forms but to look more deeply into the pragmatic aspects of discourse processing.

If there is even the slightest clue that the child is bilingual or multilingual every effort must be made to test the child in his or her strongest language(s). Some probing on this point may be necessary since it may not occur to the child, or to his parents, to tell some teacher or diagnostician, "By the way, I can read and write in Mandarin." They may not see this fact as relevant in an English speaking society or school. It may, however, be of considerable importance to an appropriate assessment of the child's actual capabilities. If a "disability" is suspected, where children are thoroughly bilingual or even multilingual, it is mandatory to assess their abilities in each of the languages they know. Usually this will involve only English and one other language, but in exceptional cases three or even more languages might be involved. To make a convincing case for a "learning disability," it is necessary to show that problems appearing in one of the child's languages also appear in the other.

There is no theory of language acquisition that will support the thesis that "learning disabilities" will only be manifested in French, or any other particular language. Deep semiotic processing problems, the kind that affect language capacity in a general way, or possibly other semiotic representational processes as well, are bound to manifest themselves in a variety of ways and cannot logically be limited to just one of a multilingual's languages. On the other hand, if problems are just apparent in one of two or more language systems a child possesses, it follows that the difficulties are likely to be within the normal range experienced by second language learners and that no real "learning disability" exists at all.

Phase four, for children identified as having special semiotic problems in more than one language or other semiotic modality, would involve a complete discourse analysis along the lines of Damico (1985a, 1985b, 1991) leading into recommendations for therapeutic intervention of an appropriate sort. At this point assessment merges with instruction (alias therapy) so completely that the two can no longer be profitably distinguished.

It would seem that procedures for intervention could benefit as much from an investigation of language instructional methods that work (cf. Oller and Richard-Amato, 1983; and Richard-Amato, 1988) as assessment of abilities and disabilities of LEPs could from the findings of language testing research. More particularly, pragmatically motivated procedures that deal with problems in the full richness and scope of normal experience will have a far better chance of success than discrete-point oriented procedures that are generally acknowledged to be recipes for failure (see Coles, 1978).

Here are a few heuristic guidelines for assessment in general. Samples of discourse, or assessment procedures themselves, should always involve performances in engaging contexts of semiotic representation. Wherever possible a variety of sources of evidence should be examined, e.g., multiple languages, dialects, kinesic representations, and sensory-motor performances. The objective should always be to find the child's optimal capabilities not to define some set of disabilities. Judgments should never be considered final but should be subject to constant updating, revision, and rechecking. No single test should form the basis for assessment. It should not be the basis for any final judgment. In the final analysis our goal is to set the child up for success, not for failure.

Notes

[1] Interestingly, Olson (1986) goes even further than Oller (1981). Subsequently, however, I believe we have followed the same river of thought (see Oller, 1989; Olson, 1986; Langer, 1987; and Sternberg, 1987).

[2] According to an unpublished study reported on at this meeting by Dr. Sherry R. Migdail, as few as 50 out of 1,000 students in a typical middle America school district were observed to have some form of genuine special education need (e.g., mental retardation, language-disorder/learning-disability, etc.). Yet, as Dr. Alba Ortiz observed in her presentation at this conference, a far higher percentage of students are misidentified as needing special education.

[3] Of course, the implication that a term like "reading" (or "listening" or even "spelling," all of which occur elsewhere in Swanson's paper) can be construed as "specific" is absurd on its face. Reading is as complex as any process known to modern science. Neither is it distinguishable except in superficial ways from all that accompanies it -- reasoning, arguing, imagining, etc. To suggest that such a process achieves the sought after "specificity" is to reveal the shallowness of thinking that characterizes the whole "social consensus" that constitutes "the field of LD".

[4] Cummins (1986) cites Rueda and Mercer (1985) who claimed that the distinction between "learning disabled" and "language disordered" for minority children is typically a matter of whether there is a "psychologist" or a "speech-pathologist" on the placement committee. Cummins concludes that the distinction is essentially arbitrary" (1986, p. 29).

[5] It should be noted that the latter authors, according to their own bibliography, only had access to summarial presentations of the pragmatic criteria they attempted to employ. Also, they compared only 12 "learning disabled" children as determined by the criteria set by the State of Alabama with 12 normals defined as such in view of their performance at "expected academic grade level". The authors apparently assume, without justification, that the children

identified by the state's criteria really are "learning disabled," but this is precisely the premise that needs to be questioned. Unless independent evidence of "learning disability" exists in those 12 children, evidence that would be missing for the "normals" against whom they are to be compared, the pragmatic criteria for evaluation cannot be tested with the experimental design that was in fact employed. In the final analysis, only some of the pragmatic criteria proposed by Damico and company did discriminate between the "disabled" and "normal" groups. However, this may be as much a consequence of group selection as of the criteria. Besides, it has been argued that significant difficulties can be expected for children that depart substantially from the norm on any one of the pragmatic criteria under consideration.

[6] Note that we do not use the term "disabled" here to legitimize it, nor do we agree that children in general to whom the label is attached are as it describes them. Our point here is to enable all children, LEPS and non-LEPs, normal and exceptional, to have access to the full range of educational benefits to which they are legitimately entitled.

[7] Krashen (1981, 1982, 1985) has argued that affective resistance to normal second language acquisition may occur in high anxiety or otherwise disturbing contexts. Assuming he is correct in this, every effort should be made to avoid the kinds of social conditions that might constitute or at least augment the mounting of such barriers.

References

Adams. C.. and Bishop, D.V.M. (1990). Conversational characteristics of children with semantic pragmatic disorder. I: Exchange structure, turntaking, repairs and cohesion. British Journal of Disorders of Communication, 24 (3), 211-239.

Alderson, J.C. (1983). The cloze procedure and proficiency in English as a second language. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 205-217). Rowley, MA: Newbury House.

Alderson, J.C. and Hughes, A., (Eds.), (1981). Issues in language testing. ELT Documents No. 111. London: British Council.

Anderson, J. (1969). Application of cloze procedure to English learned as a foreign language. Unpublished doctoral dissertation, University of New England, Australia.

Anderson, R.C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42, 145-170.

Asher, J.J. (1969). The total physical response approach to second language teaching. Modern Language Journal 53, 3-17.

Asher, J.J. and Garcia, R. (1969). The optimal age to learn a foreign language. Modern Language Journal 53, 334-341.

Asher, J.J. and Price, B.S. (1967). The learning strategy of the total physical response: some age differences. Child Development 38, 1219-1227.

Audet, L.R., and Hummel, L.J. (1990). A framework for assessment and treatment of language learning disabled children with psychiatric disorders. Topics in Language Disorders, 10 (4), 57-74.

Auerbach, E.R. (1989). Toward a social-contextual approach to family literacy. Harvard Educational Review, 59, 165-181.

Bachman, L. (1990). Fundamental considerations in language testing. New York: Oxford.

Bachman, L. and Palmer, A.S. (1983). The construct validity of the FSI Oral Interview. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 154-169). Rowley, MA: Newbury House.

Bachman, L. and Clark, J.L.D. (1987). The measurement of foreign/second language proficiency. Annals of the American Academy of Political and Social Science 490, 20-33.

Baker, E.L. (1980). Introduction. In E.L. Baker and E.S. Quellmalz (Eds.), Educational testing and evaluation: Design, analysis, and policy. Beverly Hills, CA: Sage.

Baratz, J. (1969). A bidialectal task for determining language proficiency in economically disadvantaged Negor children. Child Development 40, 889-901.

Barrett, R. (1976). A new integrative test of writing ability. Paper presented at the Annual Meeting of Region V of the National Association of Foreign Students of America, Chicago.

Beers, C.S. and Beers, J.W. (1980). Early identification of learning disabilities: facts and fallacies. Elementary School Journal 81, 67-76.

Benbow, C.P., and Arjmand, O. (1990). Predictors of high academic achievement in mathematics and science by mathematically talented students: A longitudinal study. Journal of Educational Psychology, 82 (3), 430-441.

Binet, A. (1911). New investigation upon the measure of the intellectual level among school children. L'annee psychologique, 5, 145—201. Reprinted in The development of intelligence in children (274-329). Tr. E. Kite., Baltimore: Williams and Williams.

Binet, A. and Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveaux intellectuel des anormaux. *L'anlaee psychologique* *11*, 191-24-i. Reprinted in *The development of intelligence in children* (41-88). Tr. E. Kite., Baltimore: Williams and Williams.

Bishop, D.V.M., and Adams, C. (1990). Conversational characteristics of children with semantic pragmatic disorder. II: What features lead to a judgment of inappropriacy. *British Journal of Disorders of Communication*, *24* (3), 211-239.

Bloom, A. (1987). *The closing of the American mind*. New York: Simon and Schuster.

Bloom, B.J. (1976). *Human characteristics and school learning*. New York: McGraw Hill.

Bloom, B.J. and Krathwohl, D.R. (1977). *Taxonomy of educational objectives*. New York: Longman.

Bormuth, J.R. (1970). *On the theory of achievement test items*. Chicago: University of Chicago Press.

Bower, T.G.R. (1971). The object in the world of the infant. *Scientific American* *225*, 30-38.

Bower, T.G.R. (1974). *Development in Infancy*. San Francisco: Freeman.

Boyle, J.P. (1987). Intelligence, reasoning, and language proficiency. *Modern Language Journal*, *71*, 277-88.

Briere, E. (1966). Quantity before quality in second language composition. *Language Learning* *16*, 141-151.

Briere, E. (1973). Cross-cultural biases in language testing. In J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), *Focus on the learner* (pp. 214-227). Rowley, MA: Newbury House.

Brown, H.D. (1973). Affective variables in second language acquisition. *Language Learning* *23*, (3) 1-244.

Bruner, J.S. (1984). Language, mind, and reading. In H. Goelman, A. Oberg, and F. Smith (Eds.), *Awakening to literacy* (pp. 195-201). Portsmouth, NH: Heinemann.

Burks, A.W., (Ed.), (1958). *Collected writings of C.S. Peirce*, Volumes VII and VIII. Cambridge, MA: Harvard University.

Carroll, J.B. (1961). Fundamental considerations in testing for English proficiency of foreign students. In Testing the English proficiency of foreign students (pp. 31-40). Washington, D.C.: Center for Applied Linguistics. Also in H.B. Allen and R.N. Campbell, (Eds.), Teaching English as a second language: A book of readings (pp. 313-320). New York: McGraw Hill.

Carroll, J.B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. Foreign Language Annals, 1, 131-151.

Carroll, J.B. (1983a). Psychometric theory and language testing. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 80-107). Rowley, MA: Newbury House.

Carroll, J.B. (1983b). Studying individual differences in cognitive abilities: Through and beyond factor analysis. In R. Dillon and R. Schneck, (Eds.), Individual differences in cognition (pp. 1-28). New York: Academic.

Carroll, J.B., Carton, A.S., and Wilds, C.P. (1959). An investigation of cloze items in the measurement of foreign languages. College Entrance Examination Board Research and Development Report, Harvard University. ERIC ED 021513.

Carroll, J.B., and Sapon, S.M. (1959a). Modern Language Aptitude Test (Form A). New York: Psychological Corporation.

Carroll, J.B., and Sapon, S.M. (1959b). Modern Language Aptitude Test Manual . New York: Psychological Corporation.

Cazden, C.B. (1985). The ESL teacher as advocate. Paper presented at the Nineteenth Annual TESOL Convention, New York.

Chase, A. (1977). The legacy of Malthus: The social costs of the new scientific racism. New York: Knopf.

Chihara, T., T. Sakurai, and J.W. Oller, Jr. (1989). Background and culture as factors in EFL reading comprehension. Language Testing 6 (2), 143-151.

Chomsky, N. (1956). Three models for the description of language. In I.R.E. Transactions on Information Theory, vol. IT-2, pp. 113-124. Reprinted with corrections in R.D. Luce, R. Bush, and E. Galanter, (Eds.), (1965), Readings in mathematical psychology, vol. II, New York: Wiley.

Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT.

Chomsky, N. (1972). Language and mind. New York: Harcourt, Brace, and Jovanovich.

Chomsky, N.A. (1975). Reflections on language. New York: Pantheon.

Chomsky, N. (1980a). On cognitive structures and their development. In M. Piatelli-Palmarini, (Ed.), Language and learning: The debate between Jean Piaget and Noam Chomsky (pp. 35-54). Cambridge, MA: Harvard.

Chomsky, N. (1980b). Rules and representations. New York: Columbia University.

Chomsky, N. (1988). Language and problems of knowledge. Cambridge, MA: MIT.

Church. A. (1951). The need for abstract entities in a semantic analysis. Daedalus &0, 100-112.

Clifford, Ray T. (1980). Language proficiency testing: a dynamic model. Paper presented at the Second International Language Testing Symposium, May 21-23, Darmstadt, Germany.

Cohen, A. (1980). Testing language ability in the classroom. Rowley, MA: Newbury House.

Coles, G.S. (1978). The learning disabilities test battery: Empirical and social issues. Harvard Educational Review 48, 313-340.

Condon, W.S. and Ogston, W.D. (1971). Speech and body-motion synchrony of the speaker-hearer. In D.L. Horton and J.J. Jenkins, (Eds.), The perception of language (pp. 150-173). Columbus, Ohio: Merrill.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of the research findings and explanatory hypotheses. Working Papers on Bilingualism No. 9, 1-43.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. Review of Educational Research 49, 222-251.

Cummins, J. (1981). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. Paper presented at the InterAmerican Language Proficiency Assessment Symposium, Arlie House, Virginia, March. In C. Rivera (Ed.), Communicative competence approaches to language proficiency assessment: Research and Application. Clevedon, England: Multilingual Matters.

Cummins, J. (1983a). Functional language proficiency in context: Classroom participation as an interactive process. In W. J. Tikunoff (Ed.), Compatibility of the SBIS features with other research on instruction for LEP students. San Francisco: Far West Laboratory.

Cummins, J. (1983b). Heritage language education: A literature review. Toronto, Canada: Ministry of Education.

Cummins, J. (1983c). Language proficiency and academic achievement. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 108-130). Rowley, MA: Newbury House.

Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy Clevedon, England: Multilingual Matters.

Cummins, J. (1986). Empowering minority students: A framework for intervention. Harvard Educational Review 56, 18-35.

Cummins, J. and Mulcahy, R. (1978). Orientation to language in Ukrainian-English bilingual children. Child Development 49, 1239-1242.

Cummins, J. and Swain, M. (1986). Bilingualism in education: Aspects of theory, research, and practice. London: Longman.

Davies, A. (1984). Review of J.W. Oller, Jr., (Ed.), Issues in language testing research (Newbury House, 1983). Language Testing 1, 111-114.

Damico, J.S. (1980, November). A clinical approach to discourse analysis in school age children. Mini-seminar presented at the American-Speech-Language-Hearing Association. Detroit, MI.

Damico, J.S. (1985a). Clinical discourse analysis: A functional approach to language assessment. In C.S. Simon (Ed.), Communication skill.s and classroom success (pp. 165-204). London: Taylor and Francis.

68

BEST COPY AVAILABLE

Damico, J.S. (1985b). The effectiveness of direct observation as a language assessment technique. Unpublished doctoral dissertation. University of New Mexico, Albuquerque.

Damico, J.S. (1991). Descriptive assessment of communicative ability in LEP students. In E.V. Hamayan and J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 157-218). Austin, TX: Pro-Ed.

Damico, J.S. and Oller, J.W. Jr. (1980). Pragmatic versus morphological/syntactic criteria for language referrals. Language, Speech, and Hearing Services in Schools 11, 85-94.

Damico, J.S. and Oller, J.W. Jr. (1985). Spotting language problems San Diego: Los Amigos Research Associates.

Damico, J.S., Oller, J.W., Jr., and Storey, M.E. (1983). The diagnosis of language disorders in bilingual children: pragmatic and surface-oriented criteria. Journal of Speech and Hearing Disorders 48, 285-294.

Davidson, F. (1988). An exploratory modeling survey of the trait structures of some existing language test data sets. Unpublished doctoral dissertation. University of California. Los Angeles.

Davies, A.(1984)

Duncan, S. E. and DeAvila. E. A. (1979). Bilingualism and cognition: some recent findings. NABE Journal 4, 15-50.

Ebbinghaus, H. (1897). Uber eine neue Methode sur Prufung geistiger Fahigkeiten und ihre Anwendung bei Schulkindern, Zeitschrift fur angewandte Psychologie, 13, 401-459.

Eco, U., Santambrogio, M., and Viola, P. (Eds.), (1988). Meaning and mental representations. Bloomington, IN: Indiana University Press.

Einstein, A. (1941). The common language of science. In Out of my later years (pp. 111-113). Secaucus, NJ: Citadel; also in J.W. Oller, Jr., (Ed.), (1989), Language and experience classic pragmatism (pp. 61-65). Lanham, MD: University Press of America.

Einstein, A. (1944). Remarks on Bertrand Russell's theory of knowledge. In P.A. Schilpp, (Ed.), The philosophy of Bertrand Russell (pp. 277-291). New York: Tudor Library of Living Philosophers; also in J.W. Oller, Jr., (Ed.), (1989), Language and experience: classic pragmatism (pp. 21-29). Lanham, MD: University Press of America.

ETS (1970). Manual for Peace Corps language testers. Princeton, New Jersey: Educational Testing Service.

Evola, J., E. Mamer, and B. Lentz. (1980). Discrete point versus global scoring for cohesive devices. In J.W. Oller, Jr. and Kyle Perkins (Eds.), Research in Language Testing (pp. 177-181). Rowley, MA: Newbury House.

Farhady, H. (1983a). The disjunctive fallacy between discrete-point tests and integrative tests. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 311-322). Rowley, MA: Newbury House.

Farhady, H. (1983b). New directions for ESL proficiency testing. In J. W. Oller, Jr, (Ed.), Issues in language testing research (pp. 253-269). Rowley, MA: Newbury House.

Farhady, H. (1983c). On the plausibility of the unitary language proficiency factor. In J W Oller, Jr., (Ed.), Issues in language testing research (pp. 11-28). Rowley, MA: Newbury House.

Finocchiaro, M. (1964). Teaching English as a second language. New York: Regents.

Fisch, M., et al., (Eds.), (1982). Writings of C.S. Peirce: A chronological edition. Volume 1. Indianapolis: Indiana University Press.

Fouly, K.A, Bachman, L.F., and Cziko, G.A. (1990). The divisibility of language competence: A confirmatory approach. Language Learning, 40 (1), 1-21.

Freire, P., and Macedo, D. (1987). Literacy: Reading the word and the world. S. Hadley, MA: Bergin and Garvey.

Galton, F. (1869). Hereditary genius. New York: Fontana Library.

Gardner, H. (1983). Frames of mind. New York: Basic Books.

Gardner, H. (1989). Assessment in context: The alternative to standardized testing. In B. Gifford (Ed.), Report to the Commission on Testing and Public Policy (cited by Gardner 1990 as 'forthcoming).

Gardner, H. (1990). The difficulties of school: Probable causes, possible cures. Daedalus, 119 (2), 85-113.

Gardner, H., and Hatch, T. (1989). Educational Researcher, 18 (8), 4-10.

Goodman, J.F., Freed, B., and McManus, W.J. (1990). Determining exemptions from foreign language requirements: Use of the Modern Language Aptitude Test Contemporary Educational Psychology, 15 (2), 131-141.

Goodman, K. (1967). Reading: A psycholinguistic guessing game. ,Journal of the Reading Specialist 4, 126-135.

Goodman, K. (1986). What's whole in whole language? Portsmouth, NH: Heinemann.

Goodman, K., and Goodman, Y. (1977). Learning about psycholinguistic processes by analyzing oral reading. Harvard Educational Review 47, 317-333.

Goodman, K., Goodman, Y., and Flores, B. (1979). Reading in the bilingual classroom: Literacy and biliteracy. Rosslyn, Virginia: National Clearinghouse for Bilingual Education.

Gould, S.J. (1981). The mismeasure of man. New York: W.W. Norton.

Gregg, K (1988). Epistemology without knowledge: Schwartz on Chomsky, Fodor, and Krashen. Second Language Research 4, 66-, 6.

Guilford, J.P. (1967). The nature of human intelligence. New York: McGraw Hill.

Gunnarsson, B. (1978). A look at the content similarities of between intelligence, achievement, personality and language tests. In J.W. Oller, Jr. and K Perkins, (Eds.), Language in education: testing the tests (pp. 17-35). Rowley, MA: Newbury House.

Guttman, L. (1965). The structure of interrelations among intelligence tests. In Educational Testing Service, Proceedings of the 1964 Invitational Conference on Testing Problems (pp. 25-36). Princeton, NJ: Educational Testing Service.

Hakuta, K. (1986). Mirror of language: The debate on bilingualism. New York: Basic Books.

Hakuta, K. and Diaz, R. (1984). The relationship between bilingualism and cognitive ability: a critical discussion and some new longitudinal data. In K. Nelson, (Ed.), Children's language, Vol. 5, Hillsdale, New Jersey: Lawrence Erlbaum.

Hallahan, D.P. and Cruickshank, W. (1973). Psychoeducational foundations of learning disabilities. Englewood Cliffs, New Jersey: Prentice-Hall.

Hamayan, E.V. and J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students. Austin, TX: Pro-Ed.

Hamp-Lyons, L. (1985). Two approaches to teaching reading: A classroom-based study. Reading in a Foreign Language, 3, 363-373.

Harris, D.P. (1985). Some forerunners of cloze procedure. Modern Language Journal 69 (4), 367-376.

Hartshorne, C. and Weiss, P., (Eds.), (1931-1935). Collected writings of C.S. Peirce, Volumes I-VI. Cambridge, MA: Harvard University.

Herrnstein, R.J. (1973). IQ in the meritocracy. Boston: Little Brown.

Herrnstein, R.J. and Wagner, A.R. (1981). Quantitative analysis of behavior. Cambridge, MA: Ballinger.

Hirsch, E.D. (1987). Cultural literacy. Boston: Houghton-Mimin.

Howey, KR. and Zimpher, N.L. (1989). Profiles of preservice teacher education: Inquiry into the nature of programs. Albany, NY: State University of New York.

Hughes, A. and Porter, D., (Eds.), (1983). Current developments in language testing (63-74). London: Academic.

Jakobson, R. (1980). The framework of language. Ann Arbor: University of Michigan.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? Harvard Educational Review, 39, 1-23.

Jensen, A.R. (1980). Bias in mental testing. New York: Free Press.

John-Steiner, V. (1985). Notebooks of the mind: exploration of thinking. Albuquerque, NM: University of New Mexico Press.

Kagan, D.M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks Principle. Review of Educational Research, 60 (3), 419-469.

Kaufman, A S. (1979). Intelligent testing with the WISC-R. New York: Wiley.

Kessler, C. and Quinn, M.E. (1980). Positive effects of bilingualism on science problem-solving abilities. In J.E. Alatis, (Ed.), Georgetown University Roundtable on Languages and Linguistics. Washington, D.C.: Georgetown University.

72

Kirk, S.A and Elkins, J. (1975). Characteristics of children enrolled in the Child Service Demonstration Centers. Journal of Learning Disabilities 8, 630-637.

Kirk, S.A., McCarthy, J.J., and Kirk, W.D. (1968). Illinois Test of Psycholinguistic Abilities. Urbana: University of Illinois.

Kolers, P.A. (1968). Bilingualism and information processing. Scientific American 21, 78-84.

Krashen, S.D. (1981). Second language acquisition and second language learning. Oxford: Pergamon.

Krashen, S.D. (1982). Principles and practice in second language acquisition. Oxford: Pergamon.

Krashen, S.D. (1985). The input hypothesis: issues and implications. London and New York: Longman.

Kunnan, A.J. (1990). DIF in native language and gender groups in an ESL placement test. TESOL Quarterly, 24 (4), 741-746.

Lado, R. (1961). Language testing. New York: McGraw Hill.

Lambert, W. E. (1975). Culture and language as factors in learning and education. In A. Wolfgang, (Ed.), Education of immigrant students. Toronto: Ontario Institute for Studies in Education.

Lane, H. (1984). When the mind hears. New York: Random House.

Lane, H. (1988). Educating the American Sign Language Minority of the United States: A paper prepared for the Commission on the Education of the Deaf. In S. Wilcox, (Ed.), Academic acceptance of American Sign Language: Sign Language Studies, Special Issue 59 (pp. 221-230). Silver Spring, Maryland: Linstok.

Langer, J.A. (1987). A sociocognitive perspective on literacy. In J.A. Langer (Ed.), Language, literacy, and culture: Issues of society and schooling. Norwood, NJ: Ablex.

Lotz, J. (1951). Natural and scientific languages. Daedalus 80, 87-88.

Lowe, Pardee. (1980). Language proficiency testing: a static model. Paper presented at the Second International Language Testing Symposium, May 21-23, Darmstadt, Germany.

Luria, A.R. (1959). The directive function of speech in development and dissolution. Word, 15, 341-352.

Luria, A.R. (1961). The role of speech in the regulation of normal and abnormal behavior. New York: Irvington.

Luria, A.R. (1979). Verbal regulation of behavior. In M. Cole and S. Cole (Eds.), The making of a mind: A personal account of Soviet psychology (pp. 104-119). Cambridge, MA: Harvard University Press.

Luria, A.R. and Yudovich, F.A. (1959). Speech development and the mental processes of the child. London: Staples.

Lynch, B., Davidson, F., and Henning, G. (1988). Person dimensionality in language test validation. Language Testing, 5(2), 206-219.

MacNamara, J. (1973). The cognitive strategies of language learning. In J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 57-65). Rowley, MA: Newbury House.

MacNamara, J. (1982). Names for things: A study of human learning. Cambridge, MA: MIT.

MacNamara, J. (1989). Review of Meaning and mental representations (see Eco, Santambrogio, and Viola, 1988).pp. 237. ISBN 0-253-33724-0. Journal of Language and Social Psychology, 8(5), 349-353.

McCord, J.S. and Haynes, W.O. (1988). Discourse errors in students with learning disabilities and their normally achieving peers. Journal of Learning Disabilities. 21, 237-243.

McKinney, J.D. (1984). The search for subtypes of specific learning disabilities. Journal of Learning Disabilities 17, 43-51.

McKnight, R.T. (1982). The learning disability myth in American education. Journal of Education 164, 351-359.

Mercer, J.R. (1973). Labelling the mentally retarded. Los Angeles: University of California Press.

Mercer, J.R. (1984). What is a racially and culturally non-discriminatory test? A socialistic and pluralistic perspective. In C.R. Reynolds and R.T. Brown (Eds.), Perspectives on bias in mental testing (pp. 293-356). New York: Plenum.

Miller, L. (1990). The roles of language and learning in the development of literacy. Topics on Language Disorders, 10(2), 1-24.

Moore, E.C., et al., (Eds.), (1984). Writings of C.S. Peirce: A chronological edition. Volume 2. Indianapolis: Indiana University.

Nagel, E. (1959). Charles Sanders Peirce: A prodigious but little known American philosopher. Scientific American 200, 185- 192.

Nagel, E. (1979). Teleology revisited and other essays in the philosophy and history of science. New York: Columbia University.

National Advisory Committee on Handicapped Children. (1968). Special education for handicapped children (First Annual Report). Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Newcomer, P.L. and Hammill, D.D. (1975). ITPA and academic achievement: a survey. Reading Teacher 28, 731-741.

Newmark, L. (1966). How not to interfere with language learning. International Journal of American Linguistics 32. Also in E.W. Najam and C.T Hodge. (Eds.), Language learning: The individual and the process. Bloomington, Indiana: Indiana University and in J.W. Oller, Jr. and P. Richard-Amato, (Eds.), Methods that work: A smorgasbord of ideas for language teachers (pp. 49-58). Rowley, MA: Newbury House.

Newmeyer, F. (1980). Linguistic theory in America. New York: Academic.

Nolen, S.B., and Haladyna, T.M. (1990). Personal and environmental influences on students beliefs about effective study strategies. Contemporary Educational Psychology, 15(2), 116-130.

Oller, J.W., Jr. (1970). Dictation as a device for testing foreign language proficiency. Workpapers in TESL: UCLA 4, 37-42. Also in English Language Teaching 25 (1971), 254-259.

Oller, J.W., Jr. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. Modern Language Journal 56, 151-158.

Oller, J.W., Jr. (1973). Cloze tests of second language proficiency and what they measure. Language Learning 23, 105-118.

Oller, J.W., Jr. (1979). Language tests at school. London: Longman.

Oller, J.W., Jr. (1981). Language as intelligence? Language Learning 31, 465-492.

Oller, J.W., Jr. (1983a). A consensus for the 80's? In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 351-356). Rowley, MA: Newbury House.

Oller, J.W. Jr., (Ed.) (1983b). Issues in language testing research. Rowley, MA: Newbury House.

Oller, J.W., Jr. (1986). Communication theory and testing: What and how? In C. Stansfield, (Ed.), Proceedings of the second international invitational Conference on the Test of English as a Foreign Language (pp. 99-179).Princeton, New Jersey: Educational Testing Service.

Oller, J.W., Jr (1988). The input hypothesis: Issues and implications by S.D. Krashen (1985). London and New York: Longman. Language 64, 171-173.

Oller, J.W., Jr. (Ed.), (1989). Language and Experience: Classic Pragmatism. Lanham, MD: University Press of America.

Oller, J.W., Jr. (1991). Language and bilingualism: More tests of tests. Cranbury, NJ: Bucknell University Press.

Oller, J.W.,Jr. and Conrad, C. (1971). The cloze procedure and ESL proficiency. Language Learning 21, 183-196.

Oller, J.W., Jr. and Perkins, K, (Eds.), (1978). Language in Education: Testing the Tests. Rowley, MA: Newbury House.

Oller, J.W., Jr. and Perkins, K, (Eds.), (1980). Research in Language Testing. Rowley, MA: Newbury House.

Oller, J.W., Jr., and Richard-Amato, P. (1983). Methods that work: A smorgasbord of ideas for language teachers. Rowley, MA: Newbury House.

Olson, D.R. (1986). Intelligence and literacy: The relationships between intelligence and the technologies of representation and communication. In R. Sternberg and R. Wagner (Eds.), Practical Intelligence. New York: Cambridge University.

Oltman, P.K, and Stricker, LJ. (1990). Developing homogeneous TOEFL scales by multidimensional scaling. Language Testing, 7(1), 1-12.

Oltman, P.K, Stricker, L.J., and Barrows, T.S. (1990). Analyzing test structure by multidimensional scaling. Journal of Applied Psychology, 75, 21-27.

Ortiz, A.A. and Yates, J.R. (1983). Incidence of exceptionality among hispanics: Implications for manpower planning. NABE Journal 7, 41-54.

Palmer, A.S. Groot, P.J.M. and Trosper, G.A., (Eds.).(1981). The validation of oral proficiency tests: Selected papers from The Colloquium on the Validation of Oral Proficiency Tests. Washington, D. C.: TESOL.

Peirce, C.S. (1878). How to make our ideas clear. Popular Science, Monthly, 12, 286-392. In Hartshorne and Weiss (1934), 248-271 and in Oller (1989), 53-58, 99-106.

Peirce, C.S. (1906). Truth. In Hartshorne and Weiss (1931), 3-18 and in Oller (1989), 193-201..

Piaget, Jean. (1947). The psychology of intelligence. Tottowa, New Jersey: Littlefield Adams.

Piatelli-Palmarini, M. (1980). Language and learning: The debate between Jean Piaget and Noam Chomsky. Cambridge, MA: Harvard University.

Politzer, R.L., Hoover, M.R., and Brown, D. (1974). A test in proficiency in black standard and non-standard speech. TESOL Quarterly, 8, 27-35.

Porter, D. (1983). The effect of quantity of context on the ability to make linguistic predictions: a flaw in a measure of general proficiency. In A. Hughes and D. Porter, (Eds.), Current developments in language testing (pp. 63-74). London: Academic.

Prestine, N.A. (1990). The struggle for control of teacher education: A case study. Educational Evaluation and Policy Analysis, 11(3), 285-300.

Prutting, C.A., and Kirchner, D.M. (1987). A clinical appraisal of the pragmatic aspects of language. Journal of Speech and Hearing Disorders, 52, 105-119.

Purcell, E.T. (1983). Models of pronunciation accuracy. In J. W. Oller, Jr., (Ed.), Issues in language testing research (pp. 133-153). Rowley, MA: Newbury House.

Quintero, E., and Huerta-Macias, A. (1990). All in the family: Bilingualism and biliteracy. The Reading Teacher, 44(4), 306-312.

Rand, E. (1972). Integrative and discrete-point tests at UCLA. Workpapers in TESL: UCLA 6, 67-78.

Rand, E. (1976). A factored homogeneous items approach to the UCLA ESL Placement Examination. Paper presented at the Tenth Annual TESOL Convention, New York City.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press.

Richard-Amato, P. (1988). Making it happen. New York: Longman.

Rist, Ray C. (1982). On the application of ethnographic inquiry to education: procedures and possibilities. Journal of Research in Science Teaching 19(6), 439-450.

Roid, G. H., and Haladyna, T.M. (1982). A technology for test item writing. New York: Academic Press.

Rueda, R. and Mercer, J.R. (1985). Predictive analysis of decision-making with language minority handicapped children. Paper presented at the BUENO Center Third Annual Symposium on Bilingual Education, Denver.

Russell, B. (1919). Introduction to mathematical logic. London: Allen and Unwin.

Sang, F., Schmitz, B., Vollmer, H.J., Baumert, J., and Roeder, P. M. (1986). Models of second language competence: a structural equation. Language Testing 3, 54-79.

Savignon, S. (1971). A study of the effect of training in communicative skills as part of a beginning college French course on student attitude and achievement in linguistic and communicative competence. Unpublished doctoral dissertation, University of Illinois, Champaign-Urbana, Illinois.

Savignon, S. (1983). Communicative competence: Theory and classroom practice (Texts and contexts in second language teaching). Reading, MA: Addison-Wesley.

Schumann, J. (1975). Affective factors and the problem of age in second language acquisition. Language Learning 26, 135-143.

Scott, R. (1979). Measures of Adult English Proficiency: Final report. Washington, D C.: Office of Education.

Scovel, T. (1988). A time to speak: A psycholinguistic inquiry into the critical period for human speech. Cambridge, MA: Newbury House.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics 10, 209-231.

Shepard, L.A. and Smith, M.L. (1983). An evaluation of the identification of learning disabled students in Colorado. Learning Disability Quarterly 6, 115-127.

Shuy, R.W. (1978). Problems in assessing language ability in bilingual education. in H. Lafontaine, H. Persky, and L. Golubchick, (Eds.), Bilingual education Wayne, New Jersey: Avery.

Shuy, R.W. (1981). Conditions affecting language learning and maintenance among Hispanics in the United States. NABE Journal 6, 1-18.

Simich-Dudgeon, C. (1987). Involving LEP parents as tutors in their children's education. ERIC/CLL News Bulletin 10 (2), 3-4.

Skutnabb-Kangas, T. and Toukomaa, T. (1976). Teaching migrant children's mother tongue and learning the language of the host country in the context of the socio-cultural situation of the migrant family. Helsinki: The Finnish National Commission for UNESCO.

Smith. F (1975). Twelve easy ways to make learning to read difficult, and one difficult way to make it easy. In F. Smith, (Ed.), Psycholinguistics and reading. New York: Holt, Rinehart, and Winston.

Smith, F. (1978). Understanding Reading. Second edition, New York: Holt, Rinehart, and Winston.

Smith, F. (1982). Writing and the Writer. New York: Holt, Rinehart, and Winston.

Smith, F. (1984). The creative achievement of literacy. In H. Goelman, A. Oberg, and F. Smith (Eds.), Awakening to literacy (pp. 135-142). Portsmouth, NH: Heinemann.

Smith, F. (1989). Overselling literacy. Phi Delta Kappan, 70, 352-359.

Spearman, C. (1904). "General intelligence" objectively determined and measured. American Journal of Psychology 15, 201-293.

Spearman, C. (1927). The abilities of man. New York: Macmillan.

Spolsky, B. (1968a). Language tests: The problem of validation. TESOL Quarterly 2, 88-94.

Spolsky, B. (1968b). What does it mean to know a language? Or how do you get someone to perform his competence? Paper read at the Second Conference on Problems of Foreign Language Competence, University of Southern California, Los Angeles. Also in J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 164-176). Rowley, MA: Newbury House.

Spolsky, B. (1983). Fourteen years on—later thoughts on overall language proficiency. In A. Hughes and D. Porter, (Eds.), Current developments in language testing (pp. 39-42). London: Academic.

Spolsky, B. (1985). A critical review of Krashen. Lecture presented at the University of New Mexico.

Sternberg, R.J. (1987). Second game: A school's eye view of intelligence. In J.A. Langer (Ed.), Language, literacy, and culture: Issues of society and schooling. Norwood, NJ: Ablex.

Stump, T. A. (1978). Cloze and dictation tasks as predictors of intelligence and achievement. In J.W. Oller, Jr. and K. Perkins, (Eds.), Language in education: testing the tests (pp. 36-64). Rowley, MA: Newbury House.

Swain, M., Dumas, G., and Naiman, N. (1974). Alternatives to spontaneous speech: elicited translation and imitation as indicators of second language competence. Working Papers on Bilingualism 3, 68-79.

Swanson, H.L. (1988). Toward a metatheory of learning disabilities. Journal of Learning Disabilities 21, 196-209.

Taylor, W. (1953). "Cloze procedure": A new tool for measuring readability. Journalism Quarterly 30, 415-433.

Taylor, W. (1956). Recent developments in the use of"cloze" procedure. Journalism Quarterly 33, 42-48, 99.

Taylor, W. (1957). "Cloze" readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology 41, 19-26.

Terrnan, L.M. (1925). Genetic studies of genius. Palo Alto: Stanford University.

Terrnan, L.M. and Merrill, M.A. (1960). Stanford-Binet intelligence scale; manual for the third revision, form L-M. Boston: Houghton Mifflin.

Terman, L.M. and Oden, M.H. (1947). The gifted child grows up. Palo Alto: Stanford University.

Trabue, M.R. (1916). Completion test language scales. New York: Teachers College, Columbia University.

Thurstone, L.L. (1924). The nature of intelligence. London: Paul Trench and Trubner.

Thurstone, L.L. (1938). Primary mental abilities. Chicago: University of Chicago.

Thurstone, L.L. (1947). Multiple-factor analysis: A development and expansion of the vectors of mind. Chicago: University of Chicago.

Thurstone, L.L. and Thurstone, T.G. (1941). Factorial studies of intelligence Psychometric monographs, no. 2. Chicago: University of Chicago.

Upshur, J.A. (1967). English language tests and the prediction of academic success. In National Association of Foreign Student Affairs, Selected conference papers of the Association of Teachers of English as a Second Language.

Upshur, J.A (1969a). Measurement of oral communication. In H. Schrand, (Ed.) IFS documentation, Leistungtionzenstaum fur Fremdsprachenforschung (53-80). House.

Upshur, J.A. (1969b). Productive communication testing: progress report. Paper presented at the Second International Congress on Applied Linguistics, Cambridge, England. Also in J.W. Oller, Jr. and J.C. Richards, (Eds.), (1973), Focus on the learner (pp. 177-183). Rowley, MA: Newbury House.

Upshur, J.A. (1979). Functional proficiency theory and a research role for language tests. In E.J. Briere and F.B. Hinofotis, (Eds.), Concepts in language testing: Some recent studies (pp. 75-100). Washington, D. C.: TESOL.

Upshur, J.A. and Fata, J. (Eds.), (1968). Problems in foreign language testing (Special Issue Number 3). Language Learning. Ann Arbor: University of Michigan.

Upshur, J.A and Homburg, T.J. (1983). Some relations among language tests at successive ability levels. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 188-202). Rowley, MA: Newbury House.

Valette, R. (1964). The use of dictee in the French language classroom. Modern Language Journal 39, 431-434.

Valette, R. (1967). Modern Language Testing New York: Harcourt, Brace, Jovanovich.

Vigil, N. and Oller, J.W.,Jr. (1976). Rule fossilization: A tentative model. Language Learning 26, 281-295.

Vollmer, H.J. and Sang, F. (1983). Competing hypotheses about second language ability: a plea for caution. In J.W. Oller, Jr., (Ed.), Issues in language testing research (pp. 29-79). Rowley, MA: Newbury House.

Vygotsky, L.S. (1934). Thought and language. Moscow-Leningrad: Sozekgiz. Tr. by E. Hanfmann and G. Vakar and printed in English (1962), Cambridge, MA: MIT.

Vygotsky, L.S. (1978). Mind in society: The development of higher psychological processes. Edited by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman. Cambridge, MA: Harvard University.

Walters, J., and Gardner, H. (1985). The development and education of intelligences. In F. Link (Ed.), Essays on the intellect (pp. 1-21). Washington, D.C.: Curriculum Development Associates.

Walters, J., and Gardner, H. (1986a). The crystalizing experience. In R.J. Sternberg and J.E. Davidson (Eds.), Conceptions of giftedness (pp. 306-331). New York: Cambridge, University Press.

Walters, J., and Gardner, H. (1986b). The theory of multiple intelligences: Some issues and answers. In R.J. Sternberg and R. Wagner (Eds.), Practical intelligence: Origins of competence in the everyday world (pp. 163-182). New York: Cambridge University Press.

Wells, G. (1986). The meaning makers. Portsmouth, NH: Heinemann.

Whipple, Guy Montrose. (1915). Manual of mental and physical tests. Part II. Complex processes. Second edition, revised. Baltimore, MD: Warwick and York.

Wilcox, S., (Ed.), (1988). Academic acceptance of American Sign Language: Sign Language Studies, Special Issue 59. Silver Spring, Maryland: Linstok.

Response to John Oller's Presentation

Fred Davidson
University of Illinois, Urbana/Champaign

Well, half of me wants to say sort of this is really easy because I agree. I do agree very deeply. The title of my talk is, "From the Trenches." In this paper for this meeting, John Oller has presented a thorough, theoretical and philosophical basis for motivated, proactive change in the assessment of language minority students in the United States. In my reaction, I shall do two things. I'm going to briefly summarize and interrupt his main points, give you a glimpse of the rest of that 66-page document, and then attempt to relate his philosophical stance to the pragmatic necessities of language minority students. Now, it is this second section that has caused me to title my paper, "From the Trenches." Oller's work is broad reaching and provocative. From my background as a language tester, who has worked with small and large language assessment data sets, I have decided to challenge myself and discuss how his proposal might be implemented in the front line trench battles of language testing in school setting.

First the summary: Oller's paper has three parts. In the first part, he reviews primary and non-primary language testing literature. He discusses the heritage which language testing shares with intelligence measures as well as the historical link of language testing with structural linguistics. These two trends are primarily responsible for the prevalence of discrete-point testing approaches, and the question John raises or implies many times is -- is it appropriate to consider language ability as the sum of many parts? By the second section of the paper, Oller's beliefs are clear, when on page 20, he says, "Happily a movement toward pragmatic, holistic testing is now discernable."

Much of the first section of his paper seems directed at this conclusion -- a conclusion which I share very deeply. Language ability is indeed a complex mental trait and holistic integrative testing should hold forth more than it does. Oller says near the end of his paper, "It is difficult to over estimate the pervasive influence of analytic, discrete-point thinking in the study of exceptionalities." And it is precisely this pervasive influence that I've taken as my mandate: how to expand the framework of language assessment measures in the reality of school based decision making, and I am going to return to this later.

Second, Oller has a review of relevant points from the recent history of educational measurement. He cites Roid and Haladyna, "There is a chance for endless mapping sentences, facts, and facet elements with lack of agreement among developers being a major determent to progress." And then he goes on to say," when the focus is shifted from a list of items, (which is a poor characterization in any case of any non-finite domain of sentences) to the generative basis which underlies the representations that constitute that domain, we have some hope of achieving reliability and validity." The multi-facet nature criterion reference measurement or, strictly speaking, a domain referencing, is anathema to good language testing, Oller seems to say. I agree generally with this, but I suggest that criteria can also be holistic, and I've done some work in the design and implementation criterion reference test specifications that are pragmatic and holistic. The other major component of this section is citation of the work of Gardner on multiple intelligences as that is central to part three. I want to deal with it in my discussion of that part.

In part three, Oller sketches his own model of human systems of representation, and there are three diagrams in there ending at the one that integrates Gardner with Oller. He calls this his own general semiotic capacity model. This is by far the most philosophically challenging section of the paper. My impression is that Oller is expanding on the notion of a general factor of language to encompass multiple components, and that he is utilizing Gardner's work to do so. Oller now views language as a global factor that contains components. This is very clear in the paper and this is very welcome. He closes with a series of assessment recommendations for teaching and testing language minority students. Much of this discussion centers around the nature of disabilities. He claims that the handling of language minority students has been heavily conditioned by the history of measurement, of language disorders, and/or learning disabilities. I have seen this, first hand, in my work with K through 12 ESL and bilingual data in the state of Illinois. I agree heartily.

He has several recommendations at the very end, including the use of "pragmatically motivated procedures" that deal with the problems in the full richness and scope of normal experience as well as a call for multiple measures about which I will speak specifically below. He seems to regret the difficulty of implementing change in language minority student education. The ease with which a disability or remediation paradigm can rule the day prompts him to say, "What chiefly stands in the way of the needed theoretical and practical development is the uncritical acceptance of the present social consensus, if researchers and practitioners alike are willing to acquest to the status quo of existing categories like language disorders, learning disabilities, mental retardation, and in general to the whole diagnosis and remediation paradigm the needed reform of theory and practice is bound to be delayed if it ever comes at all." He is challenging our field then to find a way to break the uncritical acceptance of the status quo, and I'd like to take up that challenge in part in the next section.

So, a voice from the trenches.

Now, the issue here, it seems to me, is that we need to get inside the head of the people that matter. All assessment is done within the context of decision making. There is a real good paper by Jack Upshur from 1970 and Lyle Bachman extends it in his 1990 textbook. The person making the decision may or may not be a test designer and, if so, may or may not subscribe to the philosophical shifts which Oller promotes, and with which I heartily agree, so this begs the question, why? What causes the acquiescence that bothers John Oller and bothers me? Let me offer a practical, real world answer. I believe that we need to legitimize the change necessary for the assessment of language minority students. This legitimization requires two components. First, full-proof argument and second, logistical ease, i.e., that the new must be as easy to implement as the old. First, full-proof argument should affect assessment score users on a philosophical strong ground as Oller has done as well as be an elegant simplicity, and I'd like to offer an example of the later. Drawing heavily upon an excellent paper in Language Testing by Mats Oscarson, 1989. I highly recommend it. Oscarson argues that if modern language teaching is more focused on the learner then the learner should be consulted in the assessment process. He argues, therefore, that language testing should include self-report. At the very beginning of his paper, Oscarson notes that there are fundamentally two types of assessment, external and internal. The former, external, is imposed from outside of the learner. Most tests are actually external. The latter are self-report of some sort or another and reflects the internal goals, agenda, and motivations of the learner, goals which may or may not match the external tests. Oscarson's paper closes with samples of self-report and language testing and the particular appropriacies of those samples to K through 12 is not really relevant here. What is at issue here is the undeniable simplicity of Oscarson's argument. The differentiation of assessment into self and non-self in my eye is equal to the philosophical paradigm shift that separated criterion referencing from norm referencing. Hudson and Lynch, Language Testing, 1984 and Glaser 1963, whom they cite, note the following about the difference between norms and criteria. They note that if achievement happens in the classroom then a normalizing test will actually unskew a curve. All teachers after they teach want people to achieve. Apply a normalizing norm referenced test to that, and you will actually convert it back to a Bell Curve. Affectively, the achievement will be statistically squashed. That's a powerful argument which appeals to teachers everywhere. I maintain that Mats Oscarson's argument, that testing needs to be internal and external, is equally simple and powerful.

Several years ago, I was fortunate to be in a seminar with John Oller at UCLA. There I presented a case for something I then called and still call "multiple referencing" which is a super-ordinate term to link criterion referencing and norm referencing and other references, as yet to be determined. I believe that self-report is actually a form of test reference, call it self-referencing, on equal stature to that of criteria and norms. Furthermore, I believe the simple elegance of Oscarson's argument elevates self-referencing to the status of norms and criteria. The simple, elegant, undeniable elevation of the new to the status to the old is one crucial component to breaking the acquiescence which Oller condemns. In this particular instance of proposed change, I believe multiple referencing is not really a new concept just a new term. Oller even appeals for it at the very end of his paper, as have many others who have used terms like multiple criteria and multiple indicators, and I have a whole scad of references here on that. I maintain that terms like multiple indicators and multiple criteria help us see multiple sources of evidence within a certain score reference, but why not attack the number of references as well, and that's why I proposed self-referencing.

But this isn't enough. An argument in favor of expansion of the number of score references, which in essence, John does at the end of his paper, is not the only necessity by far. We need to make the change work. I often pose the following question to my language testing students. Two situations; Situation A: You are an administrator at a school, a decision maker. You have 900 new international students arrive at your school, and you must decide their English proficiency. You consult a single norm referenced discrete-point test score. Situation B: You are the same person. You have 900 new international students arrive at your school, and you must decide their English proficiency. You consult a single norm referenced test score, a single criterion referenced test score, and you interview each student for self-report. The issue is that the entire technological history of logistical ease and human measurement is intertwined with the summative discrete-point test score. We cannot get away from what we appear to do so well. Clinical, detached, quasi objective discrete-point norm referenced testing, we have that down pat. A couple of years ago at a conference, I met Edward DeAvila, a developer of the LAS assessment battery. He showed me a computer expert system program to help a decision maker navigate multiple information sources, some of which constituted multiple references. As I recall, he had both norms and criteria, and I think that this program was or was a refinement of one developed for the Chicago Public School System. Now, I'm not proposing, necessarily, that a computerized expert system can automate the navigation of multiple references and a broader range of what John calls pragmatically motivated procedures, but I do claim that unless we do something to break the logistical strangle-hold of norm referenced summative discrete-point tests, we are doomed to fail. Let's hit them with both barrels. Let's use the elegant simple arguments, and let's make routine the complexity of dealing with language testing as it should be dealt with.

86

In closing, I would like to echo the sentiment of Anne Frank, "I do believe that people are basically good at heart." ...despite the way this sounds. I do agree with Anne Frank. People are basically good at heart, and this includes the staunchest decision maker/addicts of norm referenced test scores. I believe, rather, that what happens is not that they consciously reject the persuasion of Oscarson, Oller, and others, but rather that such change is felt to be logistically impossible. Let's work on that feeling.

Response to John Oller's Presentation

Myriam Met
Montgomery County Public Schools, Maryland

I agree with both John Oller and Fred Davidson. I'm just a simple practitioner, so what I'd like to do with you this morning is try to extrapolate some of the implications from foreign language practice, from the paper that John has written, and from the remarks Fred has shared with you this morning. I'd like to talk briefly about the notion of global testing of proficiency and tie that to what I think is a more important and valuable trend for all of us, which is classroom assessment of language skills.

The first part, global proficiency, I think, draws from the buzz word in the foreign language profession today (and it has been for the last decade), which is "proficiency." You might find this definition interesting because it's a somewhat different view of the term "proficiency" from the one that I was familiar with when I worked in ESL and bilingual programs (about six years ago). In foreign language proficiency, one is never "proficient." One is only proficient to perform certain tasks or language functions, in certain contexts or settings about certain topics or contents, and with a degree of both linguistic and socio-cultural accuracy. To some extent, all of us are limited proficient in that none of us, even the most ideal, (but non-existent) educated native speaker is ever completely proficient to perform all language tasks, in all contexts, in all contents, with the same degree of linguistic and socio-cultural accuracy. That is an important concept to which I'll come back in a little while when I talk about classroom proficiency and some definitions.

In the 1980s, the American Council on Teaching of Foreign Languages ACTFL undertook, along with the Educational Testing Service, to develop a global proficiency measure, which was called, not surprisingly, the ACTFL/ETS oral proficiency rating scales. What's interesting about the scales, for those of you who are not foreign languages professionals, is the fact probably, that for the first time, in the history of foreign language teaching in this century, there exists a common metric for the assessment of secondary and post-secondary students, a standardized instrument that allows everyone to agree on what the terms mean. The term "proficient," then, never really meant proficient to do everything all the time, everywhere, in every way possible, but simply to perform certain tasks in certain settings at a certain degree of accuracy as defined by the scales. That doesn't mean that everyone agrees that the scales themselves are perfect. There's no general consensus that this is the only reliable measure. In fact, there's a great deal of debate raging over the content validity of the proficiency scales. But one of the points, which is important for non-foreign language professionals to note, is that this is one instrument that everybody can focus their attention on and begin to talk about as a way of looking at student performance.

I bring that up because in a previous life, one which I enjoyed a great deal and miss a great deal, I worked with ESL and bilingual programs. One of the greatest frustrations was the lack of appropriate instruments to find out what children knew and were able to do. In proficiency testing, one is always focusing on what the learner can do, under what circumstances, and how well. Whereas, when I worked in ESL and bilingual education, I was never quite sure what the tests were really supposed to be testing. One advantage that those who work in the assessment of language minority children should have over foreign language professionals is in the area of identifying goals and objectives. The purpose for assessing English language skills should be to find out if students have acquired the English skills necessary for successful academic performance at or above grade level. In contrast, in foreign language, we very rarely know what students are going to be able to do with their language skills. We don't know the purposes to which they will put their language skills. It's really hard to figure out how to find out what children know when you really don't know what you expect them to know and be able to do in the first place. And if you really don't know what you want them to do, how do you know what to teach them? If you don't know what to teach them, it's awfully hard to decide how to test them. That should not be the case when we work with language minority students, because we are very clear about what we want them to be able to do. We want them to be able to succeed in school. John Oller has said it very well: "Language is the key to successful endeavors, especially, in the school setting." If we know what kids are supposed to be able to do, then why aren't we finding out if they can do it?

It seems to me that entry and exit decisions were based on the wrong things when I worked in ESL/Bilingual education. If you want to know whether a student can perform well academically, the first thing you probably ought to know is: "What are the demands of the academic curriculum from the language perspective?" At every grade level and in every content domain, that may differ; therefore, a student at the third grade level may need to understand this much English, speak this much English, read that much English, write that much English. It might be different for a fourth grader learning social studies or a second grader learning science. Yet, the tests we were working with all looked at students' oral production. Some of them are very discrete point, such as whether a student could discriminate between the sounds of yellow and jello. Except when we teach the concept "matter changes form," we never use the term jello in the third grade. Yet, discriminating yellow from jello was a question on the test, and whether you got to stay in the program depended on whether you understood the difference. That kind of decontextualized assessment of language seemed to be irrelevant to what we needed the kids to be able to do.

As an ESL program director, I was always terrified when the children took their language proficiency tests. Part of me really wanted them to do well, because that was what our program was all about—helping them to succeed. We wanted every child to do as well as possible. But this little voice inside of me said, "Oh, if they do well, we won't be able to help them anymore." Because no matter what the tests said, I knew that some of those children weren't quite ready to make it on their own. It seemed to me an awfully silly way to make a decision about who gets in, who stays in, and who gets out.

It's all that which brings me to my central argument.

The most promising way, then, to address the concern of the appropriate assessment language proficiency is through instructionally-based assessments, such as the ones we have been hearing about at this symposium and certainly the ones I think are relevant from my experience with foreign language immersion programs. In foreign language immersion programs, students learn content through a language in which they have limited skills. Immersion teachers are responsible for ensuring that their students achieve the objectives of the school curriculum while gaining skills in a new language. In this respect, their roles and responsibilities parallel those of teachers who work with language minority students.

For the last four and a half years, I have been involved in a project to identify the training needs of foreign language immersion teachers and to help develop training materials to meet those needs. I would be the first to say, and I really want to stress this, that foreign language immersion is not the same as ESL or bilingual education, (nor should it be), but the needs of the teachers who work in these fields are similar in that they're all engaged in teaching content in a language that is new to their students. Also, some of you who are working in the field of developmental bilingual education may find it interesting to hear some of the training issues that are involved in foreign language immersion. We have been helping teachers learn how to teach in these foreign language settings and to find out, indeed, if children are learning. In this project, I have come to believe that the teaching of language and content should be inseparable. I am going to say that again, because I think that is the most important thing I have to say today. The teaching of language and content ought be inseparable. Language is learned best through a context and a content, particularly when the aim of the language program is to enable students to be successful academically in their new language. John Oller has just told us that language is important to all educational endeavors, and that to separate language from meaning, language from thought and cognition, and from content, is to make a mockery of the business that we're all about.

Language objectives and content objectives must be tied to one another. Both sets of objectives must be considered when planning for teaching and when planning for testing. We tell our teachers that planning for testing takes place at the time that you plan for your teaching. Teachers must identify the language demands of the curriculum and plan to include means for students to gain in language as they grow in concept attainment. Anne Snow, Fred Genese, and I have suggested elsewhere a model for the integration of language objectives with the teaching of content, and visa versa, and have demonstrated how the roles of the ESL, the bilingual teacher, the mainstream teacher, and the foreign language teacher are fulfilled within that framework. I'm not going to go into that paper here, but I do want to stress the importance of teaching language through content and the importance of considering every content lesson a language lesson as well.

Teaching and testing go hand and hand. As John points out in his paper, (a point he didn't mention this morning), testing activities should be as broad as our teaching activities. In fact, planning for testing and planning for teaching need to be done at the same time. Effective foreign language immersion teachers begin to plan by first thinking about what they want students to learn. Then, when they know what they want students to learn, they've got to figure out how they're going to know it when they see it. If they know what they want children to learn and how they're going to find out if they have learned, then the next step also falls in line, which is, how you're going to get children ready to show you or to perform their knowledge. Those are the enabling activities; that's the teaching part. So, learning and teaching and testing all belong together. Good immersion teachers, then, are able to ensure that their objectives, their teaching, and their testing all fit together, because they see them as inextricably tied to one another. And, they define their objectives, teaching, and testing both in terms of content and language.

Since teaching concepts in a new language often requires that immersion teachers use visual and other concrete experiences during instruction, it follows that similar approaches are appropriate when testing students. Students should have access to materials that help them show the teacher what they know, even when they can't always tell her.

In assessing students, immersion teachers are most concerned with finding out what students have learned and allowing students to demonstrate what they have learned. The emphasis is on what students can do and do know, not on what they don't know and can't do. John told us in his paper that what we need most in our profession are integrative tests that tie teaching to learning to assessment. In foreign language immersion, classroom-based language assessments that are conducted as part of the instructional delivery system serve a number of important masters. First, they seem to be the most appropriate way of finding out whether students have the language skills needed for academic performance, precisely, because the assessment ties language to its purpose, which is content learning. These assessments are authentic in that they measure student proficiency in the real contexts in which language use occurs. They're integrated and assess the range of skills needed in the classroom for successful academic performance. These tests, in essence, have content valid[illegible]ecause, (as I heard the term used yesterday), they "test the rig[illegible] thing." Classroom-based performance assessments put the focus where it belongs, on student growth. Performance assessments such as portfolios, systematic observation, and teacher evaluations of student products and projects are effective ways to find out about student progress in relation to the objectives we've set for them. Because they're based on student performance; they show us what students can do and do know, and they compare each student to his or her last performance, they only compare students to themselves, not to some idealized and probably non-existent average student or native speaker.

Last, they're the most appropriate way of ensuring that the delivery of instruction is commensurate with the linguistic proficiency of the student at that point in time and in that content domain.

From the day to day instructional perspective, the marriage of language assessment with content assessment helps teachers, whether they're foreign language immersion teachers or those who teach language minority students, engage in a constant formative diagnostic feedback loop. In our training of foreign language immersion teachers, we emphasize the importance of surveying students' background knowledge prior to introducing a new concept. Every teacher does that but, for foreign language immersion teachers, this also means that they must know the range of the students' linguistic ability to handle the concepts. The teacher needs to know the language demands of the curriculum objectives and the extent to which special strategies, manipulatives, and concrete materials will be necessary for instructional delivery.

Immersion teachers are content teachers, but they're also language teachers. We believe that every content lesson should be a language lesson as well, and that foreign language immersion teachers need to plan as conscientiously for language growth as they do for content. In part, planning for language growth means the teacher must be continuously assessing where students are in relationship to where they ought to be and using that assessment data to identify areas where further development of language growth is needed.

It's clear, then, that as instruction progresses and as teachers observe the growth of students, a great deal of assessment data can be collected about the achievement of both content and language objectives. These data provide important information about each individual student but, in the aggregate, data from systematic observations, checklists, portfolios, and teacher-made tests also provide information about the effectiveness of the instructional program.

In conclusion, trends in foreign language teaching and testing have two major implications for the assessment of language minority children. One is, perhaps, a different definition of proficient -- a recognization that all language users, both native and non-native, are differentially proficient to perform language tasks in different settings and at varying levels of performance. None of us is completely proficient, and this definition of proficiency renders the notion of limited proficiency almost meaningless, as a system of categorizing learners. Language minority students bring with them a rich resource in their home language and culture. The label, "limited proficiency students," as John tells us in his paper, only perpetuates a deficit model of instruction and relegates ESL and bilingual education to a compensatory role. Perhaps a more useful way at looking at proficiency, as in the newer definition in foreign language, is to describe what learners can do, under what circumstances, and how well. For language minority students, defining proficiency in terms of classroom language -- the tasks, the functions, the contexts, and the contents in which they must perform -- will allow us to focus assessment measures where they belong, on academic performance. The second implication of foreign language immersion is that the teaching and testing of English in ESL bilingual programs must be integrated with the content students are to learn. If the teaching and testing of English were more intimately tied to the learning of content, we might more effectively integrate teaching, learning, and assessment. To paraphrase the late Ron Edmonds, (and I'm sure many of you have heard this before) "All children can learn. All children must learn, and all teachers must learn to teach (...and I'll throw in my paraphrase, and equitably assess) all children."